
Developmental basis of wing pattern diversity in *Heliconius* butterflies

Joseph John Hanly



Peterhouse, University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy in Zoology

August 2017

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This thesis meets the guidelines for length set by the Degree Committee for the School of Biology.

Summary

A major challenge to evolutionary developmental biology is to understand the how modifications to gene regulatory networks can lead to biological diversity. *Heliconius* butterfly wing patterns provide an excellent example of this diversity. In particular, the species *H. melpomene* and *H. erato* display wide variation in wing pattern across their ranges in Central and South America, but wherever they co-occur, they have converged on remarkably similar wing patterns due to Müllerian mimicry.

Linkage analysis of wing pattern genes has shown that in both species, there are three genomic loci that are responsible for most of the pattern variation, and that these loci are homologous. One locus, containing the transcription factor *optix*, is responsible for red pattern elements. A set of non-coding sequences linked to some of the red pattern elements have been identified. Another locus, containing the gene *WntA*, has been linked to the shape of the forewing band elements and is responsible for variation in wing pattern development in several species of lepidoptera. A third locus, responsible for yellow pattern elements, contains multiple candidate genes that may affect wing pattern development, including the gene *cortex*, which is also linked to the industrial melanism phenotype in the moth *Biston betularia*, as well as the genes *domeless* and *washout*, linked to the Bigeye mutant in *Bicyclus anynana*.

I first investigated modifications to regulatory sequence near the transcription factor *optix*, detecting a module associated with the band pattern element. I also found that for some pattern regulatory modules at *optix*, the same sequence has independently evolved the same function in *H. melpomene* and *H. erato*, in association with non-coding sequences conserved throughout the Lepidoptera.

I then investigated gene expression differences in two morphs from either side of a hybrid zone that vary only in the presence or absence of a yellow pattern element, in order to determine a role for candidate genes at the yellow pattern locus. In *H. melpomene* the gene *cortex* was upregulated in the larval wing discs of the black morph, whereas in *H. erato* it was upregulated in the larval wing discs of the yellow

morph. In pupal wings, *washout* was differentially expressed, again in the opposite pattern in the two species, suggesting the same locus is responsible for convergent pattern modification, but by a different mechanism.

Finally, I investigated the spatial transcriptomic landscape across the wings of three different heliconiine butterflies. I identified candidate factors for regulating the expression of wing patterning genes, including genes with a conserved expression profile in all three species, and others, including genes in the Wnt pathway, with markedly different profiles in each of the three species.

Each of these studies contributes to our understanding of how gene regulatory networks can be modified to create diversity: first, at the level of *cis*-regulation, second at the level of gene interaction and expression, and lastly at the level of developmental bias and constraint.

Acknowledgements

I feel very privileged to have spent most of the last five years as part of the Butterfly Genetics group. Here, I have found a set of diverse and brilliantly minded scientists and friends, who have dragged me along on their coattails through the ups and downs of the PhD. I am particularly indebted to Professor Chris Jiggins, under whose guidance and tutelage I have been able to find my confidence as a scientist. Special mention is also owed to Dr Richard Wallbank, who trained and advised me throughout this project, to Dr Simon Martin, who guided me through the genomic analyses, Dr John Davey, who (very patiently) explained to me how to use a computer, and to Sarah Barker, who supported me in the lab, both practically and morally. I started this whole process on the same day as Ana Pinharanda, and while I think I'd have just about made it to the end without her, I can't imagine having anywhere near as much fun in the process.

All of the tissue collection for this project was carried out in the insectaries at the Smithsonian Tropical Research Institute in Gamboa, Panama. This would not have been possible without the diligence and expertise of Oscar Paneso, Liz Evans, and the group of students and technicians who helped to keep the butterflies flying. I thank Lucas Brenes and Henry Arenas Castro who assisted with the tissue dissections carried out in July 2014; Chris, Richard, Markus Möst and Steven Van Belleghem, who travelled down to the Darien with me in August 2015 to hunt for butterflies; and Liz, Rachel Crisp, Tim Thurman, and everyone who kept me company during the rainy season. Owen McMillan not only welcomed me to his group in Panama, but also advised me on many aspects of this project and injected his enthusiasm into all of it.

My project was funded by the Wellcome Trust four year programme in Developmental Biology; I thank the academic and administrative staff of this programme, and of the Department and Zoology and STRI, for their support and assistance. I also thank the staff and students of Peterhouse, who always made me feel at Peterhome. I was provided with a generous bursary by the Cambridge Philosophical Society, which supported me in my final year.

I have oculocutaneous albinism. This comes with a list of visual system abnormalities about as long as your arm, and also means I can just about manage the top two rows on an eye chart. You might imagine that this could be a bit of an issue, especially for someone who spent a good chunk of the last 4 years microdissecting butterfly wings. But in reality, it has never really been a problem, and there are a few key reasons for this. First, I was given educational support, from about the age of 3 until I left school, by the Stockport Educational Service for Sensory Impairment. Megan Barley, Rowena Curley and a litany of other people did everything in their power on an almost daily basis to give me equal access to my education. They taught me how to advocate for myself, and by their example showed me how to adapt tasks that should have been impossible into tasks that I could accomplish. Second, I've been able to travel in the wake of my older sister Roisin. It always seemed like she could do anything, so it never occurred to me that I couldn't either. Lastly, I owe a lot to my friends and colleagues, both here in Cambridge and in Stockport, who have always had my back.

To my family, Eamonn, Janet, Roisin, Paddy and Niamh, I can't thank you enough for your support, laughter, strength and love.

Collaborations and publications

Chapter 2 includes sequence data provided by Owen McMillan and Camilo Salazar. Data analysis was carried out in collaboration with Megan Supple. Parts of this chapter are in preparation for publication with Jake Morris.

Some parts of my work, not included in this thesis, have been published as part of larger collaborations:

Wallbank, R.W., Baxter, S.W., Pardo-Diaz, C., **Hanly, J.J.**, Martin, S.H., Mallet, J., Dasmahapatra, K.K., Salazar, C., Joron, M., Nadeau, N. and McMillan, W.O., 2016. Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS biology*, 14(1), p.e1002353.

Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W.R., Wu, G.C., Maroja, L., Ferguson, L., **Hanly, J.J.**, Hines, H., Salazar, C., Merrill, R.M., Dowling, A., ffrench-Constant, R., Llaurens, V., Joron, M., McMillan, W.O., Jiggins, C.D. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, 534, pp.106-110.

Van Belleghem, S.M., Rastas, P., Papanicolaou, A., Martin, S.H., Arias, C.F., Supple, M.A., **Hanly, J.J.**, Mallet, J., Lewis, J.J., Hines, H.M., Ruiz, M., Salazar, C., Linares, M., Moreira, G., Jiggins, C.D., Counterman, B.A., McMillan, W.O., Papa, R. 2017. Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution*, 1, p.0052.

Abbreviations

BWA	Burrows-Wheeler Aligner
SNP	Single Nucleotide Polymorphism
Mya	Million years ago
Kb	Kilobase (1000 base pairs)
PCR	Polymerase Chain Reaction
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
GRN	Gene Regulatory Network
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeats
3'OHK	3-hydroxykynurenine
UTR	Untranslated region (can be 5' or 3')
PCP	Planar cell polarity
BGI	Beijing Genomics Institute
GFP	Green Fluorescent Protein
GLM	Generalised Linear Model
DE	Differentially expressed
GO	Gene Ontology
IGV	Interactive Genome Viewer
NCBI	National Centre for Biotechnology Information
TE	Transposable Element
Fst	Fixed statistic
PCA	Principal component analysis
NGP	Nymphalid Ground Plan
PBS	Phosphate Buffered Saline solution
Mb	Megabase (1,000,000 base pairs)
RATT	Rapid Annotation Transfer Tool.

Contents

Chapter 1: Introduction - Development and evolution of pattern and form

Development and diversity 2

A brief history of Heliconius 6

The loci under selection – optix, WntA, cortex 12

The developing wing; Lepidoptera vs Diptera 19

Summary 23

Chapter 2: Repeated co-option of ancestrally conserved regulatory sequence in the convergent evolution of mimicry patterns 33

Introduction 33

Materials and Methods 39

Results 44

Discussion 60

Chapter 3: Pattern differences across a hybrid zone: using RNA-seq to characterise molecular differences between mimetic pattern morphs

Introduction 76

Materials and Methods 86

Results 93

Discussion 118

Chapter 4: The gene regulatory landscape of the butterfly wing – a molecular dissection of the Nymphalid Ground Plan

Introduction 136

Materials and Methods 145

Results 151

Discussion 193-210

Chapter 1

Introduction

Development and evolution of pattern and form

Evolutionary developmental biology is faced by an apparent contradiction. On the one hand, biological systems are robust, and most multicellular organisms undergo developmental processes which are highly stereotyped and repeatable, even in the face of vast variation in environmental conditions. On the other hand, biological systems have a propensity to adapt and diversify to generate seemingly endless variation in form. In short, the key thesis of developmental biology is that biological systems are robust, whereas the key thesis of evolutionary biology is that biological systems are diverse.

Of course, this apparent dichotomy is false – biological systems are both robust *and* diverse. All developmental systems are the result of natural selection, and natural selection is a population genetic process (Lynch, 2007). Likewise, much of the variation within populations results from modification to developmental programs. On occasion, the distinct conceptual and methodological approaches in the two fields have been caricatured like so; In order to understand the processes of ontogenesis, developmental biologists have tended to study highly inbred animals with low phenotypic diversity as a tool for understanding the developmental consequences of deleterious mutation. Variation is viewed as background noise, and the genome is modelled as a set of discrete, large effect loci, wired together in circuits of gene regulatory networks. On the other hand, evolutionary biologists have tended to study the naturally occurring variation found within polymorphic populations or adaptive radiations as a tool for understanding the dynamics and mechanisms of evolution, and have focused on non-deleterious traits which are under selection. Under the population genetics view, the genome is sometimes mathematically modelled as an infinitely large set of loci with infinitesimal effects.

Ultimately, to gain a full understanding of the changes that occur to a trait under selection, we must study the development of those traits, and in order to fully

understand developmental processes, we must study the mechanisms by which they evolved. It has become incumbent on biologists to work towards a coherent vision of evolution and development that integrates these two paradigms.

Here, I will discuss the theoretical framework of evolutionary developmental biology which attempts to unify robustness with diversity. I will then introduce *Heliconius*, a genus of highly diverse neotropical butterflies, which offer a unique opportunity to study mechanisms of development within the context of population genetics and ecology.

Development and diversity

Multicellular organisms must proceed from a single-celled state to a more complex multicellular state, in which different cells and tissues have specialised functions. To fulfil its function, each specialised domain requires a specific complement of interacting gene products to be present, and the exclusion of other proteins. In order to coordinate this regulatory process during development, genes are organised into gene regulatory networks, hardwired genetic regulatory codes which specify which genes are to be expressed in spatial and temporal patterns (Davidson and Erwin, 2006).

GRNs are constructed of a number of key components. Each network consists of many modular DNA sequences which are capable of receiving regulatory inputs. These inputs come in the form of activating and repressing factors, proteins or RNAs that can recognise specific sequences. This results in precise transcriptional control of genes, which is dependent on the spatial and temporal arrangement of the regulatory inputs. Many GRNs have been experimentally determined by progressive overactivation and inactivation of individual components, most notably in the sea urchin (Davidson and Erwin, 2006, McIntyre et al., 2014, Oliveri and Davidson, 2004).

Animal GRNs are the product of over a billion years of natural selection. Many networks with conserved function are conserved in diverse animal lineages, for example apical specification/eye GRN, which includes regulatory interactions between the transcription factors *Pax6/ey*, *Six3/optix*, *six2/so*, *Eya* and *Dac*, which are found in association with the development of the eye or apical structures in

Drosophila, sea urchin, *Platynereis*, and the mouse (Martik and McClay, 2015, Steinmetz et al., 2010, Cvekl and Tamm, 2004). In particular, the gene *Pax6* (called *ey* - *eyeless* in *Drosophila*) maintains its homeotic function in specifying eyes, and the coding sequence and function is conserved to the extent that expression of mouse *Pax6* in *Drosophila* will rescue eye development. While the structure of networks can be highly conserved, they may accrue changes through time by modification of interactions and addition or removal of nodes (Davidson and Erwin, 2006). Additionally, the downstream effectors of networks may change greatly: when mouse *Pax6* is expressed in flies, it triggers the development of a fly eye. This conservation of GRNs through deep evolutionary time is indicative of their crucial role in developmental processes. But these networks and their effects clearly do evolve, and any modification to a GRN begins as a mutation in the germline of one individual, and must spread to fixation in a population. If such modifications cause deleterious pleiotropic effects, they will be selected against unless they are outweighed by positive selection

Modifications to protein coding sequence were the first obvious point of call in understanding variation, as they were detectable with allozyme studies (Hubby and Lewontin, 1966, Parker et al., 1998). Examples are known of protein coding variants, and at deeper time scales protein coding variation can be associated with morphological differences, for example modification of the C-terminal domain of Ubx alters its interactions with Abd-A in inhibition of leg formation in insects relative to the crustaceans (Ronshaugen et al., 2002). However, not all mutants and variants can be mapped to protein-coding differences, and even widespread populations with apparent genetic variation in morphological traits frequently appeared to have little or no detectible allozyme variation (Mashburn et al., 1978) (King and Wilson, 1975). Also, a large amount of the known protein-coding variation in humans occurs in disease states (Lek et al., 2016), illustrating how protein coding modifications are likely to have deleterious pleiotropic consequences, especially when they occur in key parts of GRNs. For example *Pax6* mutants involved in eye and neurological defects (Glaser et al., 1994), *Gli3* mutants causing polysyndactyly and cephalic malformation (Radhakrishna et al., 1997), or *MSX1* causing cleft lip and palate and tooth agenesis (van den Boogaard et al., 2000). This has led to an increased emphasis on non-coding sequence as the material of evolution (Carroll, 2000, Stern and Orgogozo, 2009).

Particular positions in a GRN may be especially prone to evolutionary change. For example, mutations affecting genes that act earlier in a GRN are more likely to generate a phenotypic effect than later genes, owing to their greater number of downstream dependencies, meaning such mutations are more likely to cause deleterious pleiotropy. An example of such a gene is *shaven baby* (*svb*), associated with larval trichome patterns in several species of *Drosophila* (Sucena et al., 2003). Expression of *svb* in an epithelial cell is enough to induce differentiation into a trichome precursor cell. Thus the regulatory region of *svb* integrates information from many upstream genes to determine whether expression should be induced and, therefore, which cells will make trichomes. This triggers a signalling cascade that is relayed to many downstream genes that are involved in making a trichome. Hence, the larval GRN includes a trichome ‘module’ that is controlled by *svb*. This node in the GRN may gain new activating or inhibiting regulatory linkages, which provides a mechanism by which modifications to a GRN can lead to evolution of form while avoiding deleterious pleiotropic effects. This may occur by both *cis*-regulatory changes to *shavenbaby* (Frankel et al., 2012), or by post-transcriptional regulation by *mir-92a* (Arif et al., 2013). Having said this GRNs can be modified by simply adding or removing individual nodes to create novel functions. Lactase persistence, the ability to metabolise lactase beyond infancy in humans, is caused by the gain of an Oct-1 transcription factor binding site in an enhancer region of the lactase gene, causing temporal persistence of the expression of this gene (Lewinsky et al., 2005). Also, a whole GRN can be turned off wholesale, causing pleiotropic effects which are selectively advantageous, as with cave albinism in *Astyanax* cave fish (Protas et al., 2006).

While we have a good understanding of gene regulation in some systems, and a couple of very well-understood cases where there is variation in gene regulation, we still have a paucity of described examples of non-coding variants that have been under selection. GePheBase, a database of genotype-phenotype relationships, lists 1697 entries as of July 2017; of these 198 (12%) are non-coding changes related to morphology, and 31 of these (1.8% of the total) claim evidence of selection (Martin and Orgogozo, 2013). These known cases are generally not well understood in their broader developmental context. There is a need to gain an understanding of how

modifications to gene regulatory networks and developmental programs can lead to biological diversity

Appropriate study systems must be selected in order to study diversity in a framework of developmental robustness, and to understand the functional basis of diversity.

Adaptive radiations can contain a huge array of closely related species and forms which are descended from a recent common ancestor, giving them several useful key features. Many adaptive radiations are rich with measurable variation. Classic examples include cichlid fishes of the African Great Lakes, (Brawand et al., 2014) and Darwin's Finches (Grant, 2003), where reproductive isolation has arisen through the occupation of new environmental niches and evolution of adaptations in morphological characters such as beak and jaw shape, as well as pigmentation and colour pattern. As there is a recent shared common ancestor, the variation in different populations has evolved with the same or similar developmental and genomic constraints, and in recent radiations the signals of selection and adaptation in the genome are more likely to be detectable by association. Much of the variation may occur in species or populations which are interfertile, meaning QTL analysis can be performed (as long as the reproductive cycle is amenable to it).

There are also a number of features that are very useful for developmental studies, including the ability to perform manipulative experiments, for example with the experimental genetic tools available in *Drosophila*. Also, in order to perform effective linkage or association analyses, to perform genetic manipulation experiments, or gene expression analyses, it is useful to have good genomic resources. Recent technological advancement has greatly increased our ability to both perform manipulative experiments and to generate genomic data efficiently and effectively. This means that developmental studies of form and morphology are no longer restricted to traditional model systems, but can now be performed in systems which were previously not accessible to experimental perturbation, such as adaptive radiations. *Heliconius*, a genus of neotropical butterflies, is a prime example of such a group.

A brief history of Heliconius

For almost 150 years, *Heliconius* butterflies have occupied the minds of evolutionary biologists. On travelling along the Amazon River and observing the incredible diversity and variation in form, Henry Walter Bates took particular note of how the wing patterns of butterflies would periodically change as he travelled deeper inland, “as if at the touch of an enchanter's wand” (Bates, 1862). In observing *Heliconius* and other animals, he recognised that brightly coloured individuals could be distasteful or toxic to predators, that the bright and bold colour patterns were acting as aposematic cues which could be learnt by predators, and that this could lead to mimicry. Study of mimicry in *Heliconius* was furthered by Fritz Müller, who realised that aposematic populations can gain fitness through mimicry, as the increased occurrence of the aposematic signal in the populations increases the effectiveness of the signal, and the burden of teaching it is spread over a larger number of individuals. He provided a mathematical model for this frequency dependent selection, the first of its kind (Müller, 1879).

All species of *Heliconius* lay eggs on Passifloraceae, the passion vines, and while feeding upon them as larvae acquire cyanogenic compounds which make them distasteful to predators, meaning that *Heliconius* are frequently involved in Müllerian mimicry. Adult *Heliconius*, like most butterflies, use the nectar of flowers as their food source. Unlike most other butterflies, *Heliconius* also collect and digest pollen, which provides a protein-rich diet. This has consequences for life history, including a long life span and a large brain – *Heliconius* have the second largest mushroom bodies of any insect that has been studied, behind only *Apis mellifera* (Montgomery et al., 2016) and in the wild may live as long as 6 months (Ehrlich and Gilbert, 1973).

There are over 40 species of *Heliconius* distributed from southern Brazil up to the southern United States. Some of these species are monomorphic, but several of them have up to tens of different pattern morphs, and engage in rampant mimicry both with other species of *Heliconius* and with other butterflies, as well as pericopine and geometrid moths (Figure 1.1). Occasionally, this mimicry is so faithful as to have been undetectable by pattern and morphology alone; there are populations of *H. timareta* on the eastern slopes of the Andes that mimic *H. melpomene* so closely that they were only discovered by DNA sequencing. Several groups participate in mimicry

rings; for example, wherever *H. melpomene* and *H. erato* co-occur, they always mimic each others' wing pattern. These two species diverged around ten million years ago and probably evolved their wing patterns independently. Generally their patterns are closely mimetic but distinguishable to the eye when in hand, if not in flight. Even so, their mimicry has occasionally led to serious confusion; one of the type specimens used by Linnaeus to describe *H. melpomene* (which he named *Papilio melpomene*) is actually an *H. erato* (Mallet, 2015).

Heliconius melpomene and *H. erato* get their English common name “Postman butterflies” because the forms found in Trinidad – with a broad red forewing band on a black background – shared their colour palate with the uniform of the Trinidadian postal service. A name made all the more fitting by the propensity of *Heliconius* butterflies to engage in traplining behaviour; flying the same route every morning and stopping off at their preferred flowers as they go, like a postman on his rounds (Murawski and Gilbert, 1986). The Postman wing pattern, which in many areas also includes the hindwing yellow bar, can be found in several species of tropical butterfly along the Caribbean and Atlantic Coasts of Central and South America, as well as along the western Andes. As with all *Heliconius* wing patterns, these bold strokes of red, black and yellow serve as an aposematic warning of distastefulness to avian predators, and are mutually used by multiple sympatric populations in a classic example of Müllerian mimicry (Merrill et al., 2015).



Figure 1.1: First row: *H. burneyi huebneri*, *H. aoede auca*, and *H. xanthocles zamora*; second row: *H. timareta timareta* f. *timareta*, *H. doris doris*, and *H. demeter ucayalensis*; third row: *H. melpomene malleti*, *H. egeria homogena*, and *H. erato emma*; fourth row: *H. elevatus pseudocupidineus*, *Eueides heliconioides eanes*, and *E. tales calathus*; and bottom: *Chetone phyleis*, a pericopine moth. Butterflies figured are from the Neukirchen Collection, McGuire Centre, Florida. The butterflies are from populations in both Ecuador and Peru. This figure is reproduced from Wallbank et al (2015).

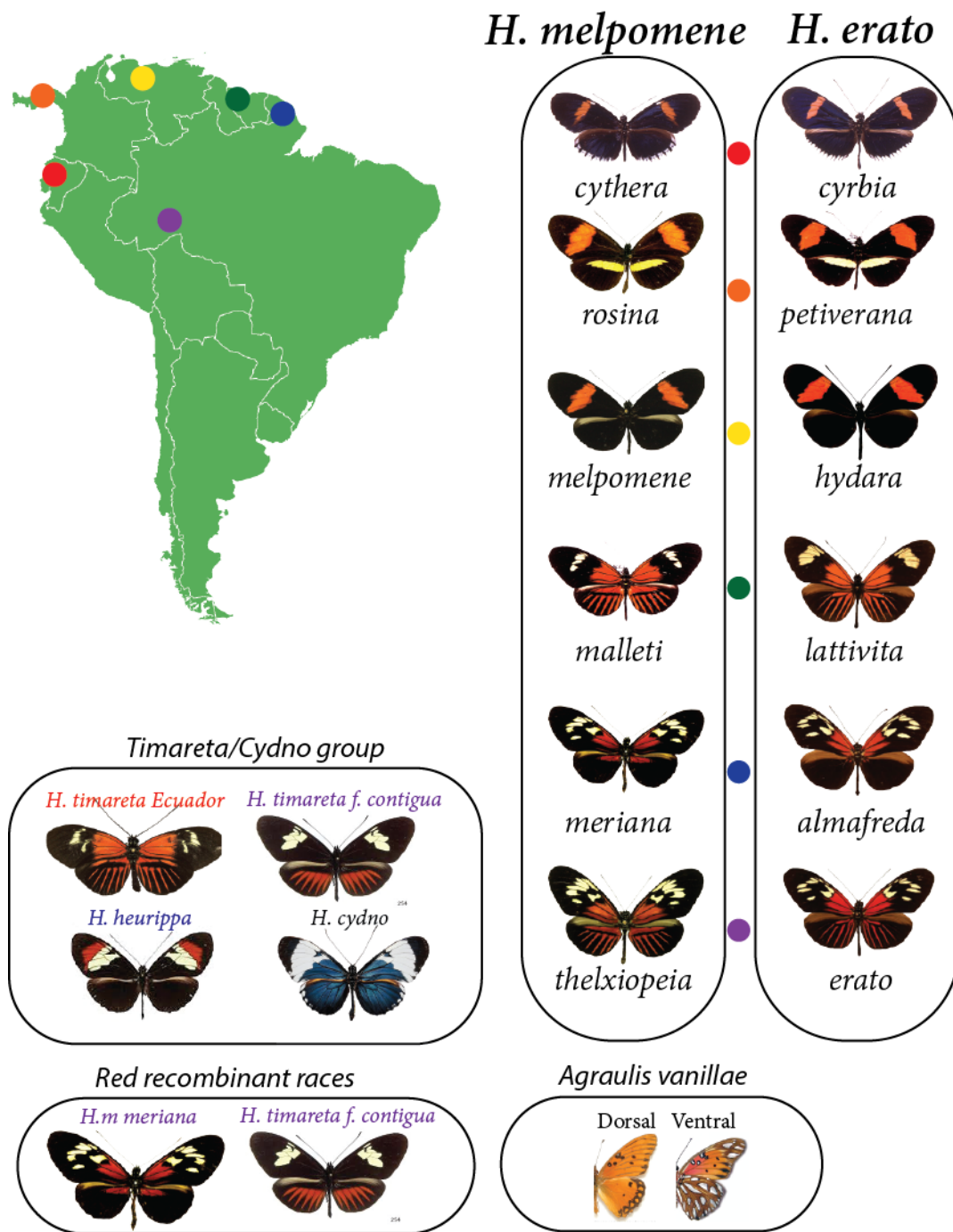


Figure 1.2 Distributions of selected mimetic races of *H. melpomene* and *H. erato*. Coloured dots indicate region of origin. A number of key additional races and species are also depicted.

From the 1950s, an understanding of the genetics of *Heliconous* wing patterns began to emerge. William Beebe and Jocelyn Crane and John Turner, who worked at the New York Zoological Society field station in Trinidad, were able to import *H. melpomene* and *H. erato* with different wing patterns from Suriname, French Guiana and Venezuela, and crossed them with each other and the local Trinidadian forms. They quickly determined that patterns were inherited as Mendelian loci (Beebe, 1955). Crane and Turner then established that there were several autosomal loci of major effect and began to determine linkage (Turner, 1973, Turner and Crane, 1962)

Crane had one particular male butterfly with an interesting wing pattern, a recombinant form (similar to *H. m. meriana* in figure 1.2) who mated with several females and started three important mapping broods. She thought he was a menace to the females, and so named him Dennis the Menace. Henceforth, the red wing pattern element of Dennis the butterfly became known as the “Dennis” patch. Thus began a long and often tortuous history of wing pattern genetic nomenclature for *Heliconius* butterflies.

Research in Trinidad stopped in December of 1964, at which point Turner returned to the UK with his stocks packed into an insulated picnic hamper, and continued his crosses in the greenhouses of Paul Sheppard at the University of Liverpool. Turner, Sheppard and others continued these crosses of the species *H. melpomene* and *H. erato* until Sheppard died in 1978, and the work was eventually published in 1985 (Sheppard et al., 1985). This paper, which includes hundreds of broods that provide information about much of the pattern variation in both *H. melpomene* and *H. erato*, describes or re-describes around 30 different colour pattern loci. Remarkably, the advent of physical linkage mapping with AFLP markers revealed that the colour pattern loci were mainly linked into three groups, and that each of the physical positions that correspond to these groups of loci in *H. melpomene* and *H. erato* are homologous (Joron 2006). This is in spite of the fact that Sheppard et al recognised that there were a number of clear differences between segregation and interaction of pattern elements in the two species, for example in the inheritance of the forewing band pattern elements (Table 1.1), which was previously assumed to indicate differences in the mechanism of wing pattern development.




















Region	<i>H. melpomene</i>		<i>H. erato</i>	
Chr 1	<i>K</i>		?	
Chr 10 - <i>WntA</i>	<i>Ac</i>		<i>Sd</i>	
			<i>St</i>	
			<i>Ly</i>	
Chr 15 - <i>cortex</i>	<i>Yb</i>		<i>Cr</i>	
	<i>Sb</i>			
	<i>N</i>			
Chr 18 - <i>optix</i>	<i>dennis</i>		<i>D</i>	
	<i>ray</i>		<i>R</i>	
	<i>band</i>		<i>Y</i>	

Table 1.1: Mimetic pattern loci in *H. melpomene* and *H. erato*. Each homologous pattern linked chromosomal region is indicated, as well as each pattern locus associated with that region. A region on chromosome 1 is linked to the *K* locus, associated with the switch between white and yellow patterns in *H. melpomene* and closely related species. The region on chromosome 10 that includes the gene *WntA* contains one locus in *H. melpomene*, *Ac*, which has alternative haplotypes causing different patterns. In *H. erato*, there are three loci in this region, each linked to different forewing band phenotypes. Similarly a region on chromosome 15, which contains the gene *cortex*, contains three distinct loci in *H. melpomene*, but only one in *H. erato*, which codes for alternative haplotypes. The region on chromosome 18, which includes the gene *optix* contains three loci in each species. The *Y* locus in *H. erato* has two alternate haplotypes, red or yellow.

Contemporaneously to the work of Turner and Sheppard in the UK, Larry Gilbert was performing his own analyses of *Heliconius* wing patterning genetics and evolution at the University of Texas (Gilbert, 2003). He allowed crosses and hybridisations to occur and observed the hybrid offspring. He then correlated his results to developmental theory – making many predictions which have been vindicated in the years since. One aspect of his approach was fundamentally at odds to the approach to workers in the UK – while Turner, Crane and Sheppard sampled the wild diversity within species of *Heliconius* and performed selective crosses with alleles of interest, Gilbert allowed hybridisation to happen between pattern forms and species essentially at random (an approach he refers to as the creation of ‘hybrid swarms’). This allowed him to observe a large array of possible variation, allelic interactions, and pattern mutants, in his words “prospecting” for interesting patterns, in particular rare recombinants that either fitted or challenged his paradigm. The resulting model will be discussed in Chapter 4. Gilbert and his students, particularly Jim Mallet, combined their understanding of the genetics of wing pattern with ecological field experiments to build a coherent narrative of the evolution and diversification of *Heliconius*, including hybrid zones, interspecific and intraspecific hybridisation, mating and foraging behaviours and host-plant interactions (Jiggins, 2017).

The loci under selection – *optix*, *WntA*, *cortex*

Linkage analysis of red pattern elements pointed to a region on chromosome 18 in both species. Initially, the gene *kinesin* seemed like the most likely candidate gene at the locus (Baxter et al., 2010), but later the gene *optix*, a short, GC-rich single-exon gene neighbouring *kinesin*, was successfully amplified (Reed et al., 2011). This gene, a transcription factor and homolog of the eumetazoan apical patterning gene Six3, was a much clearer candidate for involvement in patterning, and expression analysis by *in situ* hybridisation and then by immunofluorescence showed that the expression domains of *optix* in pupal wings closely prefigured the regions that would become red pattern elements (figure 1.3).

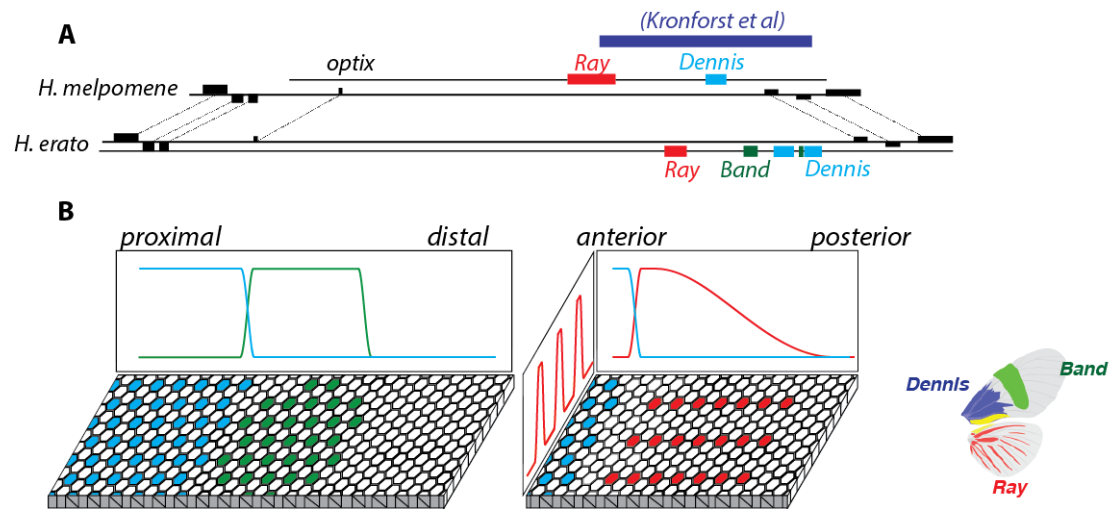


Figure 1.3: Schematic representation of *optix* locus, with functions of regulatory modules. A shows the *optix* locus as published. A series of homologous genes are described, including *optix* which sits within a large gene desert. Wallbank et al identified modules associated with *dennis* and *ray* in *H. melpomene*, and Van Belleghem et al identified modules associated with all three red pattern elements in *H. erato*. B is a cartoon of the approximate expression profiles induced by *optix* by each pattern element. The graphs indicate expression level, while the hexagonal grid indicates the wing epithelium; coloured cells are scale cells which express *optix*.

Functional verification of *optix* in the form of inhibition, knockout or overexpression (i.e. anything further than correlative expression) has not yet been achieved. However for another locus, we now have several lines of strong functional data. Multiple forewing shape alleles were mapped to a locus on chromosome 10, again in both species. This region contained the Wnt signalling pathway ligand *WntA*. (Martin et al., 2012) The same gene has also been mapped to pattern variation in one other species, *Limenitis arthemis* (Gallant et al., 2014). Expression of this gene as determined by *in situ* hybridisation recapitulates some of the pattern boundaries in wings, especially in the forewing pattern elements. Initial functional data came from injection of Wnt pathway agonist heparin and the antagonist dextran sulfate (Martin and Reed, 2014). Agonising the Wnt pathway caused loss of yellow forewing pattern elements, and antagonising it caused increase in the size of pattern elements.

Implementation of the CRISPR-cas9 system in Lepidoptera has allowed knockouts of *WntA* in tens of different species, including in multiple species of *Heliconius*. This has included both G₀ mosaic knockouts and the generation of a stable mutant line in *H. sara* (Martin et al, in submission). The effects of knockouts closely mirror that of pharmacological inhibition, causing the expansion of forewing pattern elements. Remarkably, knockout of *WntA* in the co-mimics *H. melpomene* and *H. erato* has substantially different effects on wing pattern (figure 1.4). Most notably, in *H. erato*, *WntA* appears to be an inhibitor of *optix*, but this does not appear to be the case in *H. melpomene* (Concha and Wallbank, unpublished).

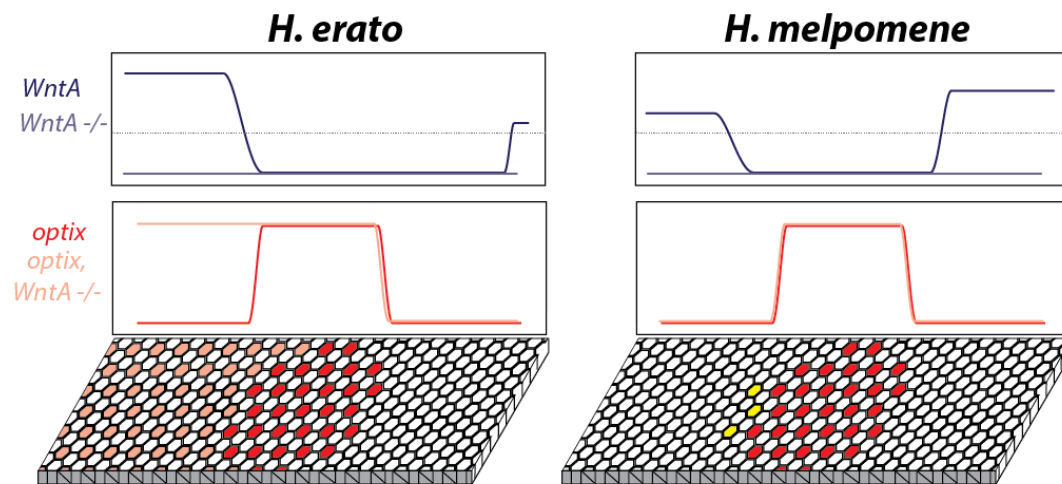


Figure 1.4 Cartoon of *WntA* function in pupal wings of *H. melpomene* and *H. erato*. The top panels show expression of *WntA*, as inferred from *in situ* hybridisations (Martin et al., 2012). The middle panel shows wild type *optix* expression (Reed et al., 2012; Martin et al., 2014), as well as inferred *optix* expression profile in *WntA* knockouts in each species. The wing epithelium cartoon indicates the effect on wing pattern in the *WntA* mutant butterflies.

The locus on chromosome 15, linked to white/yellow colour patterns, is currently the least-well understood. Nadeau et al found that pattern-associated variation occurred around the gene *cortex*, showed differential expression, as well as evidence that splice variation was present in different pattern forms (Nadeau et al., 2016). Additionally, it was found that a TE insertion in the first intron of *cortex* is linked to the *carbonaria* pattern form of the Peppered Moth *Biston betularia*, causing the industrial melanism phenotype (Van't Hof et al., 2016). There are other genes at this locus that could potentially be linked to wing pattern, including *domeless* and *washout* – these genes neighbour a non-coding region linked to the *Bigeye* mutation in *Bicyclus anynana* and have been implicated in shaping eyespot patterns in CRISPR mutagenesis experiments, and are within an interval of around 100kb which also includes *cortex* (Beldade et al 2007; Saenko et al 2010, unpublished data). It has also been shown that homozygous *Bigeye* mutants are embryonic lethal and cause segmentation defects, leading to the conclusion that *domeless* or *washout* are likely to be regulated by *engrailed* or another segmentation cascade gene (*en* has been linked to eyespot development).

Gilbert also described the pigments and scale structures that constitute *Heliconius* wing patterns (Gilbert et al., 1988). Butterfly wings are covered in coloured scales. Each scale contains a complement of pigments and ultrastructural features that contribute a specific appearance to the wing. *Heliconius* wings have three main types of scale; type I scales, which either contain 3OHK pigment making them yellow or no pigment making them structurally white, type II scales which are black and contain melanin, which can be modified to matte, gloss, or iridescent blue by ultrastructural variation, or type III scales which contain red ommochrome pigments which can vary from bright oranges to browns by interactions with pigment modifying proteins. Colour modifications to type I, II and III scales are inherited as traits unlinked to pattern variation (Kronforst et al., 2006). Structure and pigment are strongly coupled in scale cells; for example, scales with type I structure are never black or red, and intermediate scales are not found as regular features of pattern elements. Fade effects or intermediate colours like pink are achieved by interspersing scales of different types, in much the same way that pixels construct colours on computer screens. The coupling of colour and structure was further illustrated in wounding experiments (Janssen et al., 2001). The fundamental question of how butterfly wings are patterned

can thus be distilled down to a question of how scale cells get their identity, which can be viewed as a cell fate decision. In order to gain an appreciation for how this might occur, it becomes necessary to consider the developmental process of wing formation in insects.

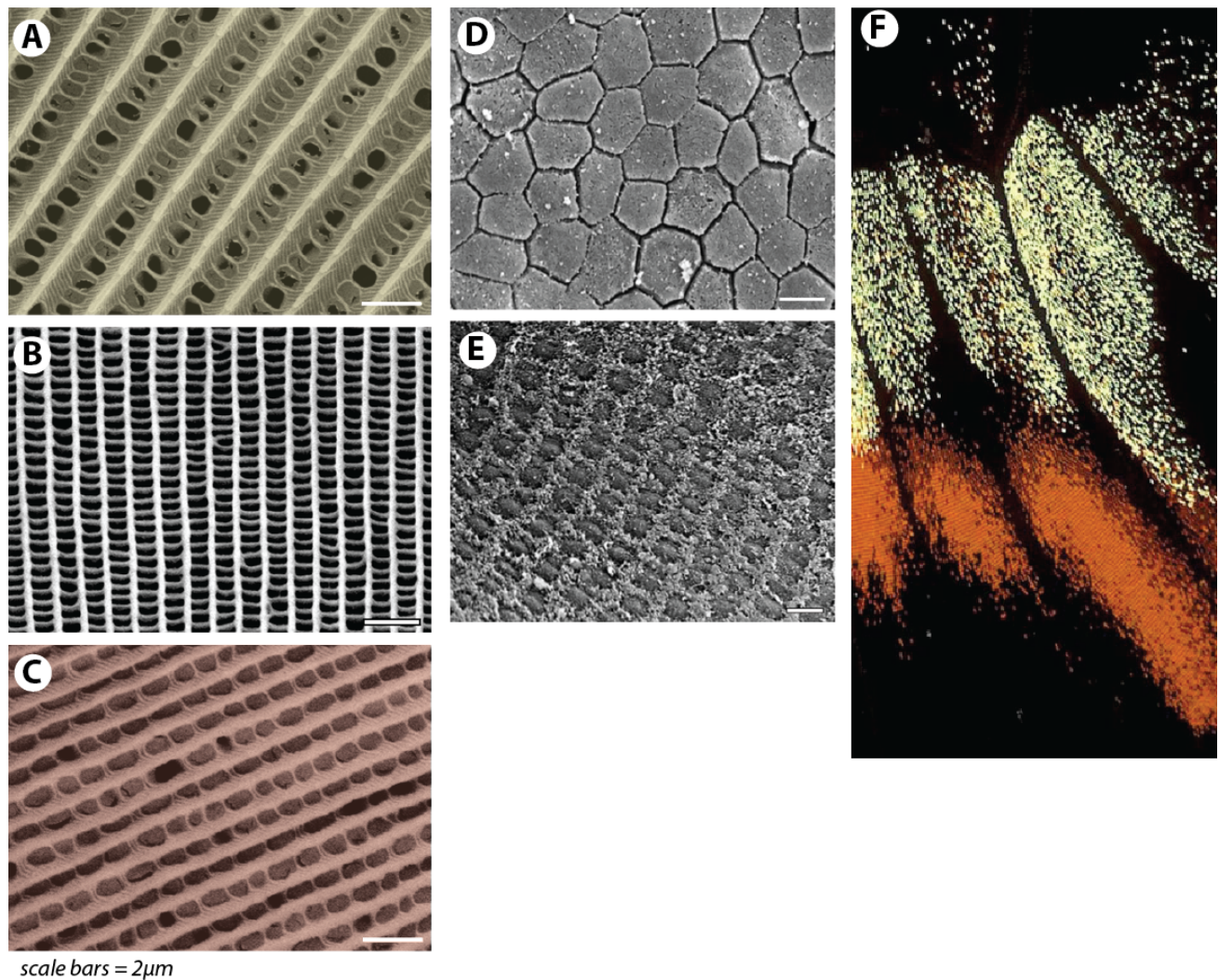


Figure 1.5: Electron micrographs indicating structure and development of scales.

A-C show the different ultrastructural arrangements of chitin on Type I (yellow/white), Type II (black) and Type III (red) scales. D shows the structure of the late larval epithelium, which consists of a typical field of undifferentiated epithelial cells, in a hexagonal arrangement (as drawn in cartoon figures 1.3, 1.4 and 1.6). E shows the pupal wing disc after the first day of pupal development; scale precursor cells are clearly differentiated from surrounding epithelial cells, and are arranged in rows. F shows a close-up of an adult butterfly wing. No intermediates between scale types are found – variations in colour are mainly achieved by intermixing the three core scale types like pixels.

The developing wing; Lepidoptera vs Diptera

Investigations of wing development in *Drosophila* have used both candidate gene approaches and mutagenesis screens to build up a wing gene regulatory network. While there are likely several key differences to the development of butterfly wings (including the Dipteran modification of T3 wings into halteres), the knowledge built in *Drosophila* can be used to guide our understanding of butterfly wing development, and a handful of genes have known Lepidopteran expression profiles which correlate with those of their *Drosophila* orthologs. These factors, present in the wing during development, may be capable of influencing the development of pattern.

A number of genes with important homeotic functions known from *Drosophila* are conserved in their expression domains. Correlative expression studies have largely been carried out in *Junonia coenia* (formerly *Precis coenia*) unless otherwise stated. For example, *Ubx* is expressed in hindwings but not forewings (Warren et al., 1994), and rare homeotic mutants have been described in which hindwings contain patches of forewing pattern, which are caused by loss of *Ubx* expression in patches of hindwing (Weatherbee et al., 1999), indicating that *Ubx* specifies the identity of the hindwing in Lepidoptera. *Apterous* (*Ap*) is expressed on the dorsal wing surface but not the ventral wing surface (Carroll et al., 1994) and in *Bicyclus*, ventralisation of the dorsal wing pattern has been linked to small voids in *Ap* expression at presumptive eyespot foci, indicating that *Ap* specifies dorsal identity in butterflies (Prakash and Monteiro, 2017). *Invected* (*inv*) was shown to be expressed in the posterior portion of the hindwing with a boundary between the R2 and M1 veins. *Scalloped* (*sd*) is expressed in all cells of the wing disc (Carroll et al., 1994). The genes *wg* and *cut* are expressed along the wing margin and their expression is correlated with adult wing shape in several genera (Macdonald et al., 2010). The gene *hedgehog* (*hh*) is expressed in the posterior compartment of the hindwing, *cubitus interruptus* (*ci*) is expressed in the anterior compartment of the hindwing, and *patched* (*ptc*) is expressed just-anterior of the A-P boundary. This correlates with their expression domains in *Drosophila*, where they contribute to specifying the anterior-posterior compartments and the corresponding boundary (Keys et al., 1999).

Genes related to epithelial identity on the wing have also been described. The butterfly homolog of the *achaete-scute* complex was found to be expressed in scale precursor cells during early pupal development (Galant et al., 1998). This corroborates the hypothesis that lepidopteran scales are homologous to *Drosophila* sensory bristles, which was based on cytological observations made in *Ephesia* moths (Stossberg, 1938). Additionally, observation of *Notch* expression in *Heliconius* indicates that scale precursors differentiate from their surrounding epithelia in a process of lateral inhibition (Reed, 2004).

Until recently, manipulation of gene expression has been difficult in the Lepidoptera as RNAi-based systems did not reliably produce experimental results (Terenius et al., 2011), and so limited genetic manipulation of gene expression has occurred. One exception to this is the evidence that RNAi against *wingless* (*wg*) causes reduced wing size in *Bicyclus anynana*, mirroring the role of *wg* in *Drosophila* (Ozsu et al., 2017). Other data about gene expression in lepidopteran wings is correlative, and relies on a candidate gene approach which has frequently depended on the existence of cross-reactive antibodies produced in other experimental systems. This mainly correlative, candidate gene data based on inference from flies leaves the possibility that there may be critical factors expressed in the wings of butterflies, or specifically in *Heliconius*, which have not been identified simply because they are either not expressed or not functional in *Drosophila*.

Of these genes, only the expression profile of the gene *Notch* has been shown in *Heliconius*. If we assume we can incorporate all of this information into our understanding of butterfly wing development, then we have built a skeleton of a butterfly wing GRN, which can feed into wing patterning. Late 5th instar larval wing discs consist of a basal lamina, with an epithelial monolayer on both the dorsal and ventral surface (wing disc eversion occurs much earlier in Lepidoptera than in Diptera). At this stage, several genes are differentially expressed along the dorsal-ventral axis of the forewing and the anteroposterior axis of the hindwing (figure 1.6A). In the hours after pupation, scale precursor cells acquire their identity by *Notch*-mediated lateral inhibition, creating evenly spaced, neatly aligned rows. They undergo two rounds of division, the first into a scale lineage and a neural lineage. The neural precursor undergoes apoptosis, and the pre-scale cell divides again into a

socket cell and a scale cell. This process requires asymmetric cell division, so the epithelial surface has acquired planar cell polarity by this time (Figure 1.6B). By around 60h of pupal development, the scales which will acquire Type I identity (i.e. yellow/white scales) begin to undergo laminar extension before their Type II and III neighbours. Also around this time the gene *optix* is differentially expressed in presumptive red scale cells (Figure 1.6C). By 100h of pupal development, all scale cells are undergoing actin-dependent laminar extension and depositing cuticle, including scale type-specific gene expression of pigment synthesis pathways (Figure 1.6D), ultimately resulting in a complete wing pattern at eclosion.

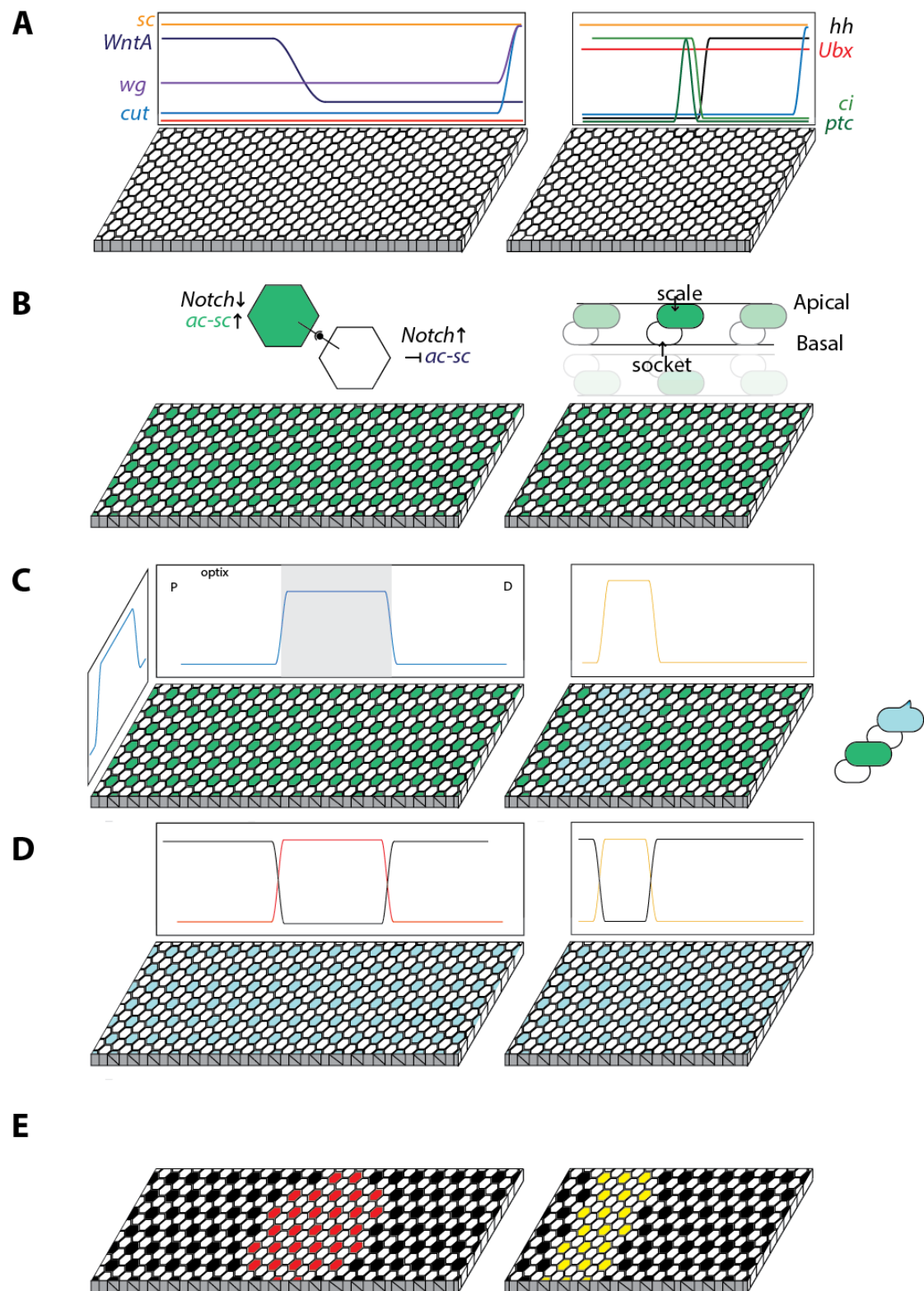


Figure 1.6: Cartoon of temporal changes in development of wing pattern. A depicts late larval wings. The expression domains of several transcription factors and signalling molecules are known at this stage in butterflies. The wings consist of parallel sheets of undifferentiated epithelia. B: During the first 18h of pupal development, scale precursor cells differentiate by a process of lateral inhibition.

These cells then undergo a series of divisions and apoptosis events that result in a scale cell and a socket cell. This process requires the presence of planar cell polarity in the tissue. C; after around 60h of pupal development, the gene *optix* is expressed in a patterned way in the wing. Also, presumptive white/yellow scale cells begin to differentiate, undergoing convergent extension. D; in mid to late pupal development, scale cell differentiation occurs across the wing, including morphological differentiation and pigment synthesis and deposition. E; this results in the adult wing pattern at eclosion.

We have some clear gaps in our knowledge of wing development in *Heliconius*, and wing pattern development generally. As well as the fact that most data is correlative, we do not know when scale precursor cells are competent to respond to cues that lead to their differentiation. We know that wounding and transplants can affect wing pattern in early pupal development, (Janssen et al., 2001) and that some of the cues that are linked to wing pattern have patterned expression in 5th instar larvae, but in order for a scale precursor to utilise any of these differentially expressed inputs, they must be expressed during a time in which the scale cell is competent to integrate this information into a coordinated output. We do not understand the physical mechanisms by which gene expression in the larval and pupal wings feeds into the wing patterns. There is evidence that in the case of *optix*, *cis*-regulatory modules are at play, but protein-DNA interactions that are capable of interfering with wing pattern have not yet been described. To resolve this, we need gene expression data for butterfly wings that is not biased by a candidate gene approach, and we need to find ways of unpicking gene regulatory networks without the arsenal of tools available to drosophilists. This will allow us to understand how butterfly wing patterns form, and how they can be modified by selection in the context of the great diversity of wing patterns in nature.

Summary

Understanding of wing pattern genetics and development in *Heliconius* has progressed greatly in the last decade in lockstep with progress in sequencing technologies and latterly in genome editing methods. This has occurred in the broader context of research in these butterflies which has allowed for wide-ranging studies of their life history and evolution, including predation, host plant interactions, vision and learning, reproductive barriers and hybridisation in speciation, mating behaviour, pheromones and adaptation to climate (Merrill et al., 2015, Jiggins, 2017). This makes *Heliconius* highly amenable to the integrated study of evolutionary processes.

In this thesis, I investigate modifications to regulatory sequence near the transcription factor *optix* so that we now have regulatory modules for all three red mimetic pattern elements in both *H. melpomene* and *H. erato*, allowing for the investigation of the roles of positional conservation and homology in convergent evolution. I then study gene expression differences in two morphs from either side of a hybrid zone that vary

only in the presence or absence of the yellow hindwing bar, in order to determine a role for candidate genes at the yellow pattern locus. In *H. melpomene* the gene *cortex* was upregulated in the larval wing discs of the black morph, whereas in *H. erato* it was upregulated in the larval wing discs of the yellow morph. In pupal wings, *washout* was differentially expressed, again in the opposite pattern in the two species, suggesting that we have identified multiple genes at the same locus, which are responsible for convergent pattern modification, but by a different mechanism, prompting further questions about how this locus can integrate patterning signals and effect scale cell identity. Finally, I screened the spatial transcriptomic landscape across the wings of three different heliconiine butterflies. I identified candidate factors for regulating the expression of wing patterning genes, including genes with a conserved expression profile in all three species, and others, including genes in the Wnt pathway, with markedly different profiles in each of the three species. This gives a view of the gene regulatory network in the wing that avoids a candidate gene-driven approach, and will hopefully inform future investigations of wing patterning in *Heliconius* and in butterflies generally.

Each of these studies contributes to our understanding of how gene regulatory networks can be modified by natural selection in the generation of diversity: first, at the level of *cis*-regulation, second at the level of gene interaction, and lastly at the level of developmental bias and constraint. Generally, the huge diversity in butterfly wing patterns gain novel shapes, features and functions in each lineage, but they do this on a backdrop of a robust and deeply developmentally conserved structure. It is my hope that this thesis will contribute to our understanding of the interface between developmental robustness and evolutionary diversity in the context of butterfly wing patterns.

References

- ARIF, S., MURAT, S., ALMUDI, I., NUNES, M. D., BORTOLAMIOL-BECET, D., MCGREGOR, N. S., CURRIE, J. M., HUGHES, H., RONSHAUGEN, M., SUCENA, E., LAI, E. C., SCHLOTTERER, C. & MCGREGOR, A. P. 2013. Evolution of mir-92a underlies natural morphological variation in *Drosophila melanogaster*. *Curr Biol*, 23, 523-8.
- BATES, H. W. 1862. XXXII. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London*, 23, 495-566.
- BAXTER, S. W., NADEAU, N. J., MAROJA, L. S., WILKINSON, P., COUNTERMAN, B. A., DAWSON, A., BELTRAN, M., PEREZ-ESPONA, S., CHAMBERLAIN, N., FERGUSON, L., CLARK, R., DAVIDSON, C., GLITHERO, R., MALLET, J., MCMILLAN, W. O., KRONFORST, M., JORON, M., FFRENCH-CONSTANT, R. H. & JIGGINS, C. D. 2010. Genomic hotspots for adaptation: the population genetics of Mullerian mimicry in the *Heliconius melpomene* clade. *PLoS Genet*, 6, e1000794.
- BEEBE, W. 1955. Polymorphism in reared broods of *Heliconius* butterflies from Surinam and Trinidad. *Zoologica*, 40, 139-143.
- BRAWAND, D., WAGNER, C. E., LI, Y. I., MALINSKY, M., KELLER, I., FAN, S., SIMAKOV, O., NG, A. Y., LIM, Z. W., BEZAULT, E., TURNER-MAIER, J., JOHNSON, J., ALCAZAR, R., NOH, H. J., RUSSELL, P., AKEN, B., ALFOLDI, J., AMEMIYA, C., AZZOUZI, N., BAROILLER, J. F., BARLOY-HUBLER, F., BERLIN, A., BLOOMQUIST, R., CARLETON, K. L., CONTE, M. A., D'COTTA, H., ESHEL, O., GAFFNEY, L., GALIBERT, F., GANTE, H. F., GNERRE, S., GREUTER, L., GUYON, R., HADDAD, N. S., HAERTY, W., HARRIS, R. M., HOFMANN, H. A., HOURLIER, T., HULATA, G., JAFFE, D. B., LARA, M., LEE, A. P., MACCALLUM, I., MWAIKO, S., NIKAIDO, M., NISHIHARA, H., OZOUF-COSTAZ, C., PENMAN, D. J., PRZYBYLSKI, D., RAKOTOMANGA, M., RENN, S. C. P., RIBEIRO, F. J., RON, M., SALZBURGER, W., SANCHEZ-PULIDO, L., SANTOS, M. E., SEARLE, S., SHARPE, T., SWOFFORD, R., TAN, F. J., WILLIAMS, L., YOUNG, S., YIN, S., OKADA, N., KOCHER, T. D., MISKA, E. A., LANDER, E. S., VENKATESH, B., FERNALD, R. D., MEYER, A., PONTING, C. P., STREELMAN, J. T., LINDBLAD-TOH, K., SEEHAUSEN, O. & DI PALMA, F. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513, 375-381.
- CARROLL, S. B. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell*, 101, 577-80.

- CARROLL, S. B., GATES, J., KEYS, D. N., PADDOCK, S. W., PANGANIBAN, G. E., SELEGUE, J. E. & WILLIAMS, J. A. 1994. Pattern formation and eyespot determination in butterfly wings. *Science*, 265, 109-14.
- CVEKL, A. & TAMM, E. R. 2004. Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. *Bioessays*, 26, 374-86.
- DAVIDSON, E. H. & ERWIN, D. H. 2006. Gene regulatory networks and the evolution of animal body plans. *Science*, 311, 796-800.
- EHRlich, P. R. & GILBERT, L. E. 1973. Population structure and dynamics of the tropical butterfly *Heliconius ethilla*. *Biotropica*, 69-82.
- FRANKEL, N., WANG, S. & STERN, D. L. 2012. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc Natl Acad Sci U S A*, 109, 20975-9.
- GALANT, R., SKEATH, J. B., PADDOCK, S., LEWIS, D. L. & CARROLL, S. B. 1998. Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. *Curr Biol*, 8, 807-13.
- GALLANT, J. R., IMHOFF, V. E., MARTIN, A., SAVAGE, W. K., CHAMBERLAIN, N. L., POTE, B. L., PETERSON, C., SMITH, G. E., EVANS, B., REED, R. D., KRONFORST, M. R. & MULLEN, S. P. 2014. Ancient homology underlies adaptive mimetic diversity across butterflies. *Nat Commun*, 5, 4817.
- GILBERT, L. 2003. Adaptive novelty through introgression in *Heliconius* wing patterns: evidence for shared genetic “tool box” from synthetic hybrid zones and a theory of diversification. *Ecology and evolution taking flight: butterflies as model systems*, 281-318.
- GILBERT, L., FORREST, H. S., SCHULTZ, T. D. & HARVEY, D. J. 1988. Correlations of ultrastructure and pigmentation suggest how genes control development of wing scales of *Heliconius* butterflies. *J. Res. Lepid*, 26, 141-160.
- GLASER, T., JEPEAL, L., EDWARDS, J. G., YOUNG, S. R., FAVOR, J. & MAAS, R. L. 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet*, 7, 463-71.
- GRANT, B. R. 2003. Evolution in Darwin's finches: a review of a study on Isla Daphne Major in the Galapagos archipelago. *Zoology (Jena)*, 106, 255-9.

- HUBBY, J. L. & LEWONTIN, R. C. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 54, 577-94.
- JANSSEN, J. M., MONTEIRO, A. & BRAKEFIELD, P. M. 2001. Correlations between scale structure and pigmentation in butterfly wings. *Evol Dev*, 3, 415-23.
- JIGGINS, C. D. 2017. *The Ecology and Evolution of Heliconius Butterflies: A Passion for Diversity*, Oxford University Press.
- KAYE, P. M., PATEL, N. K. & BLACKWELL, J. M. 1988. Acquisition of cell-mediated immunity to *Leishmania*. II. LSH gene regulation of accessory cell function. *Immunology*, 65, 17-22.
- KEYS, D. N., LEWIS, D. L., SELEGUE, J. E., PEARSON, B. J., GOODRICH, L. V., JOHNSON, R. L., GATES, J., SCOTT, M. P. & CARROLL, S. B. 1999. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science*, 283, 532-4.
- KING, M. C. & WILSON, A. C. 1975. Evolution at two levels in humans and chimpanzees. *Science*, 188, 107-16.
- KRONFORST, M. R., YOUNG, L. G., KAPAN, D. D., MCNEELY, C., O'NEILL, R. J. & GILBERT, L. E. 2006. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc Natl Acad Sci U S A*, 103, 6575-80.
- LEK, M., KARCZEWSKI, K. J., MINIKEL, E. V., SAMOCHA, K. E., BANKS, E., FENNEL, T., O'DONNELL-LURIA, A. H., WARE, J. S., HILL, A. J., CUMMINGS, B. B., TUKIAINEN, T., BIRNBAUM, D. P., KOSMICKI, J. A., DUNCAN, L. E., ESTRADA, K., ZHAO, F., ZOU, J., PIERCE-HOFFMAN, E., BERGHOUT, J., COOPER, D. N., DEFLAUX, N., DEPRISTO, M., DO, R., FLANNICK, J., FROMER, M., GAUTHIER, L., GOLDSTEIN, J., GUPTA, N., HOWRIGAN, D., KIEZUN, A., KURKI, M. I., MOONSHINE, A. L., NATARAJAN, P., OROZCO, L., PELOSO, G. M., POPLIN, R., RIVAS, M. A., RUANO-RUBIO, V., ROSE, S. A., RUDERFER, D. M., SHAKIR, K., STENSON, P. D., STEVENS, C., THOMAS, B. P., TIAO, G., TUSIE-LUNA, M. T., WEISBURD, B., WON, H. H., YU, D., ALTSHULER, D. M., ARDISSINO, D., BOEHNKE, M., DANESH, J., DONNELLY, S., ELOSUA, R., FLOREZ, J. C., GABRIEL, S. B., GETZ, G., GLATT, S. J., HULTMAN, C. M., KATHIRESAN, S., LAAKSO, M., MCCARROLL, S., MCCARTHY, M. I., MCGOVERN, D., MCPHERSON, R., NEALE, B. M., PALOTIE, A., PURCELL, S. M., SALEHEEN, D., SCHARF, J. M., SKLAR, P., SULLIVAN, P. F., TUOMILEHTO, J., TSUANG, M. T., WATKINS, H. C., WILSON, J. G., DALY, M. J., MACARTHUR, D. G. & EXOME AGGREGATION, C. 2016.

- Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536, 285-91.
- LEWINSKY, R. H., JENSEN, T. G., MOLLER, J., STENSBALLE, A., OLSEN, J. & TROELSEN, J. T. 2005. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet*, 14, 3945-53.
- LYNCH, M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci U S A*, 104 Suppl 1, 8597-604.
- MACDONALD, W. P., MARTIN, A. & REED, R. D. 2010. Butterfly wings shaped by a molecular cookie cutter: evolutionary radiation of lepidopteran wing shapes associated with a derived Cut/wingless wing margin boundary system. *Evol Dev*, 12, 296-304.
- MALLET, J. 2015. Is the type of *Heliconius melpomene* really a *Heliconius erato*?
- MARTIK, M. L. & MCCLAY, D. R. 2015. Deployment of a retinal determination gene network drives directed cell migration in the sea urchin embryo. *Elife*, 4.
- MARTIN, A. & ORGOGOZO, V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*, 67, 1235-50.
- MARTIN, A., PAPA, R., NADEAU, N. J., HILL, R. I., COUNTERMAN, B. A., HALDER, G., JIGGINS, C. D., KRONFORST, M. R., LONG, A. D., MCMILLAN, W. O. & REED, R. D. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proc Natl Acad Sci U S A*, 109, 12632-7.
- MARTIN, A. & REED, R. D. 2014. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Dev Biol*, 395, 367-78.
- MASHBURN, S. J., SHARITZ, R. R. & SMITH, M. H. 1978. Genetic Variation among *Typha* Populations of the Southeastern United States. *Evolution*, 32, 681-685.
- MCINTYRE, D. C., LYONS, D. C., MARTIK, M. & MCCLAY, D. R. 2014. Branching out: origins of the sea urchin larval skeleton in development and evolution. *Genesis*, 52, 173-85.
- MERRILL, R. M., DASMAHAPATRA, K. K., DAVEY, J. W., DELL'AGLIO, D. D., HANLY, J. J., HUBER, B., JIGGINS, C. D., JORON, M., KOZAK, K. M., LLAURENS, V., MARTIN, S. H., MONTGOMERY, S. H., MORRIS, J., NADEAU, N. J., PINHARANDA, A. L., ROSSER, N., THOMPSON, M. J., VANJARI, S., WALLBANK, R. W. & YU, Q. 2015. The diversification of

- Heliconius butterflies: what have we learned in 150 years? *J Evol Biol*, 28, 1417-38.
- MONTGOMERY, S. H., MERRILL, R. M. & OTT, S. R. 2016. Brain composition in Heliconius butterflies, posteclosion growth and experience-dependent neuropil plasticity. *J Comp Neurol*, 524, 1747-69.
- MÜLLER, F. 1879. Ituna and Thyridia: a remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. Lond*, 1879, 20-29.
- MURAWSKI, D. & GILBERT, L. 1986. Pollen flow in Psiguria warscewiczii: a comparison of Heliconius butterflies and hummingbirds. *Oecologia*, 68, 161-167.
- NADEAU, N. J., PARDO-DIAZ, C., WHIBLEY, A., SUPPLE, M. A., SAENKO, S. V., WALLBANK, R. W., WU, G. C., MAROJA, L., FERGUSON, L., HANLY, J. J., HINES, H., SALAZAR, C., MERRILL, R. M., DOWLING, A. J., FFRENCH-CONSTANT, R. H., LLAURENS, V., JORON, M., MCMILLAN, W. O. & JIGGINS, C. D. 2016. The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, 534, 106-10.
- OLIVERI, P. & DAVIDSON, E. H. 2004. Gene regulatory network analysis in sea urchin embryos. *Methods Cell Biol*, 74, 775-94.
- OZSU, N., CHAN, Q. Y., CHEN, B., GUPTA, M. D. & MONTEIRO, A. 2017. Wingless is a positive regulator of eyespot color patterns in Bicyclus anynana butterflies. *Dev Biol*.
- PARKER, P. G., SNOW, A. A., SCHUG, M. D., BOOTON, G. C. & FUERST, P. A. 1998. What molecules can tell us about populations: choosing and using a molecular marker. *Ecology*, 79, 361-382.
- PRAKASH, A. & MONTEIRO, A. 2017. apterous A Specifies Dorsal Wing Patterns And Sexual Traits In Butterflies. *bioRxiv*, 131011.
- PROTAS, M. E., HERSEY, C., KOCHANNEK, D., ZHOU, Y., WILKENS, H., JEFFERY, W. R., ZON, L. I., BOROWSKY, R. & TABIN, C. J. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet*, 38, 107-11.
- RADHAKRISHNA, U., WILD, A., GRZESCHIK, K. H. & ANTONARAKIS, S. E. 1997. Mutation in GLI3 in postaxial polydactyly type A. *Nat Genet*, 17, 269-71.
- REED, R. D. 2004. Evidence for Notch-mediated lateral inhibition in organizing butterfly wing scales. *Dev Genes Evol*, 214, 43-6.

- REED, R. D., PAPA, R., MARTIN, A., HINES, H. M., COUNTERMAN, B. A., PARDO-DIAZ, C., JIGGINS, C. D., CHAMBERLAIN, N. L., KRONFORST, M. R., CHEN, R., HALDER, G., NIJHOUT, H. F. & MCMILLAN, W. O. 2011. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, 333, 1137-41.
- RONSHAUGEN, M., MCGINNIS, N. & MCGINNIS, W. 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature*, 415, 914-7.
- SHEPPARD, P., TURNER, J., BROWN, K., BENSON, W. & SINGER, M. 1985. Genetics and the evolution of Mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 433-610.
- STEINMETZ, P. R., URBACH, R., POSNIEN, N., ERIKSSON, J., KOSTYUCHENKO, R. P., BRENA, C., GUY, K., AKAM, M., BUCHER, G. & ARENDT, D. 2010. Six3 demarcates the anterior-most developing brain region in bilaterian animals. *Evodevo*, 1, 14.
- STERN, D. L. & ORGOGOZO, V. 2009. Is genetic evolution predictable? *Science*, 323, 746-51.
- STOSSBERG, M. 1938. Die Zellvorgänge bei der Entwicklung der Flügelschuppen von *Ephestia kühniella* Z. *Zeitschrift für Morphologie und Ökologie der Tiere*, 34, 173-206.
- SUCENA, E., DELON, I., JONES, I., PAYRE, F. & STERN, D. L. 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*, 424, 935-8.
- TERENIUS, O., PAPANICOLAOU, A., GARBUTT, J. S., ELEFThERIANOS, I., HUVENNE, H., KANGINAKUDRU, S., ALBRECHTSEN, M., AN, C., AYMERIC, J. L., BARTHEL, A., BEBAS, P., BITRA, K., BRAVO, A., CHEVALIER, F., COLLINGE, D. P., CRAVA, C. M., DE MAAGD, R. A., DUVIC, B., ERLANDSON, M., FAYE, I., FELFOLDI, G., FUJIWARA, H., FUTAHASHI, R., GANDHE, A. S., GATEHOUSE, H. S., GATEHOUSE, L. N., GIEBULTOWICZ, J. M., GOMEZ, I., GRIMMELIKHUIJZEN, C. J., GROOT, A. T., HAUSER, F., HECKEL, D. G., HEGEDUS, D. D., HRYCAJ, S., HUANG, L., HULL, J. J., IATROU, K., IGA, M., KANOST, M. R., KOTWICA, J., LI, C., LI, J., LIU, J., LUNDMARK, M., MATSUMOTO, S., MEYERING-VOS, M., MILLICHAP, P. J., MONTEIRO, A., MRINAL, N., NIIMI, T., NOWARA, D., OHNISHI, A., OOSTRA, V., OZAKI, K., PAKONSTANTINO, M., POPADIC, A., RAJAM, M. V., SAENKO, S., SIMPSON, R. M., SOBERON, M., STRAND, M. R., TOMITA, S., TOPRAK, U., WANG, P., WEE, C. W., WHYARD, S., ZHANG, W., NAGARAJU, J., FRENCH-CONSTANT, R. H., HERRERO, S., GORDON, K., SWEVERS, L. & SMAGGHE, G. 2011. RNA interference in Lepidoptera:

an overview of successful and unsuccessful studies and implications for experimental design. *J Insect Physiol*, 57, 231-45.

- TURNER, J. Genetics of Race Formation in Mimetic Butterflies-is gene unit of selection. GENETICS, 1973. GENETICS SOC AM 9650 ROCKVILLE AVE, BETHESDA, MD 20814 USA, S281-S281.
- TURNER, J. & CRANE, J. 1962. The genetics of some polymorphic forms of the butterflies *Heliconius melpomene* Linnaeus and *Heliconius erato* Linnaeus. I. Major Genes. *Zoologica*, 47, 141-152.
- VAN DEN BOOGAARD, M. J., DORLAND, M., BEEMER, F. A. & VAN AMSTEL, H. K. 2000. MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat Genet*, 24, 342-3.
- VAN'T HOF, A. E., CAMPAGNE, P., RIGDEN, D. J., YUNG, C. J., LINGLEY, J., QUAIL, M. A., HALL, N., DARBY, A. C. & SACCHERI, I. J. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, 534, 102-5.
- WARREN, R. W., NAGY, L., SELEGUE, J., GATES, J. & CARROLL, S. 1994. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature*, 372, 458-61.
- WEATHERBEE, S. D., NIJHOUT, H. F., GRUNERT, L. W., HALDER, G., GALANT, R., SELEGUE, J. & CARROLL, S. 1999. Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr Biol*, 9, 109-15.

Chapter 2

Repeated co-option of ancestrally conserved regulatory sequence in the convergent evolution of mimicry patterns

ABSTRACT

Changes to *cis*-regulatory sequences play an important role in the evolution of adaptation and diversity. *Heliconius melpomene* and *H. erato* are Müllerian co-mimics, and have frequently converged on the same wing patterns in different regions. It has previously been shown that there are three homologous genomic loci that are responsible for most of the pattern variation in these species. One locus, containing the transcription factor *optix*, is responsible for red pattern elements. A set of non-coding sequences linked to some of the red pattern elements have been identified. I investigated modifications to regulatory sequence near *optix*, detecting a module associated with the *band* pattern element, and found that this module had been shared between populations by introgression. I also found that for some pattern regulatory modules at *optix*, as well as at other colour pattern loci *WntA* and *cortex*, the same sequences have independently evolved the same function in both lineages, in association with non-coding sequences conserved throughout the Lepidoptera, suggesting the presence of *cis*-regulatory hotspots of evolution.

I: INTRODUCTION

Convergence has been widely described in nature at many taxonomic levels. Examples include echolocation in whales and bats (Li et al., 2010), melanism patterns and bristle distribution in *Drosophila* (Prud'homme et al., 2006, Rogers et al., 2013, Stern and Frankel, 2013), repeated acquisition of pollination syndromes amongst the flowering plants (Stebbins, 1970) and mimicry in butterflies and other insects. Convergence is generally driven by adaptation to similar conditions, and may also be facilitated by developmental bias, whereby a constrained morphospace can restrict or direct the evolutionary trajectory of parallel lineages (Brakefield, 2006). An emerging pattern in studies of convergence is that similar changes in phenotype

often result from similar changes in genotype. Notably, homologous genes or regulatory modules are commonly used in multiple lineages to create convergent phenotypic effects. Such ‘hotspots’ have been observed to involve repeated gains and losses of function, both in coding and non-coding sequence. In some cases, single-function genes represent the only plausible solution to a particular problem. Likely examples include enzymes involved in catalytic processes like detoxification and insecticide resistance (Zhen et al., 2012, Dobler et al., 2012), and lactase persistence (Tishkoff et al., 2007, Lewinsky et al., 2005). Unlike one-step catalytic processes, developmental processes involve networks of interacting genes, proteins and physical interactions. It has been hypothesized that hotspot genes are more malleable to evolutionary change due to their central position in regulatory networks, which permits the avoidance of pleiotropy (Martin and Orgogozo, 2013, Stern and Orgogozo, 2008). Such genes have been termed ‘input-output’ loci.

For example, loss of function in regulatory modules explains convergent patterns of loss of larval trichomes in different species of *Drosophila*, caused by loss-of-function mutations in homologous conserved motifs (Frankel et al., 2012). This convergent loss of the same regulatory modules echoes similar loss of function in protein-coding genes, such as melanin synthesis genes in cave-albinism in *Astyanax* fishes (Protas et al., 2006). But while null mutations to *oca2* cause systemic and syndromic effects in the development of the cave fish, loss of specific regulatory modules of *ovo/svb* only affect larval trichome pattern and not the other primary function of this gene, which is the survival and maturation of female germ cells.

Similar patterns can be seen in cases of convergent constructive evolution, for example melanisation patterns in *Drosophila*. Within *D. melanogaster*, convergence in sexually dimorphic abdomen melanisation has been linked to the regulatory region of the gene *bab* (Rogers et al., 2013). Between species, notable examples of convergence include gain of melanic wing spots (Prud'homme et al., 2006) where convergent change in phenotype is linked to changes in the expression profile of the *yellow* gene but caused by changes to different regulatory loci; in one case in the first intron, and in the other in the 5' enhancer region.

Aside from *Drosophila*, another well-studied example of convergent and constructive cis-regulatory evolution is lactase persistence in humans – the gain of the ability to metabolize lactate into adulthood. One European and three African persistence alleles have arisen separately, all caused by single-base mutation to a *cis*-regulatory module that increases affinity for the transcription factor Oct-1, causing the expression of lactase to persist into adulthood. (Tishkoff et al., 2007, Lewinsky et al., 2005) (Jensen et al., 2011).

We are thus presented with two contrasting models for constructive convergence in regulatory sequence. In the case of melanic spots in *Drosophila*, each lineage has acquired a different, uncorrelated change that causes the same alteration to the expression profile of the regulated gene. In contrast, in the lactase persistence and trichome examples, the convergence in expression profile is driven by modifications to the same conserved ancestral element. So convergent evolution can either be driven by mutations in different regulatory modules, or to the same regulatory modules.

Heliconius wing patterns are a good example of constructive regulatory evolution and phenotypic convergence driven by Müllerian mimicry. Several species have converged on a set of geographically distinct wing patterns. The species *H. erato* and *H. melpomene* diverged around 10 mya (Kozak et al., 2015) and are mimetic in their wing pattern wherever they co-occur. This within-species divergence, combined with between-species convergence presents an opportunity to investigate the developmental basis of convergence in natural populations (Merrill et al., 2015).

As discussed in chapter 1, wing pattern variation in *Heliconius* results from developmental switches that control not just pigment synthesis but also differences in gross scale cell shape and ultrastructural differences likely driven by actin dynamics during scale development (Gilbert et al., 1988, Janssen et al., 2001, Dinwiddie et al., 2014). As such red, yellow and black pattern elements are constructed of different scale cell types (Gilbert et al., 1988).

Population association studies in both *H. melpomene* and *H. erato* have located causative sites for red wing patterns to a block of non-coding sequence downstream of the transcription factor *optix*. In *H. erato* a 65 kb region was identified (Supple et al.,

2013), while in *H. melpomene*, recombination breakpoint analysis has identified separate regions for individual wing pattern elements, a 7 kb region associated with the dennis patch (*dennis*) and a 37 kb region linked with the hindwing rays (*ray*) (Wallbank et al., 2016). The 65 kb region in *H. erato* contains conserved sequence with *H. melpomene* (Supple et al., 2013). These regions are hypothesized to contain non-coding regulatory sequences that drive expression of the *optix* gene in order to switch on red patches in different wing regions. A schematic representation of the *optix* locus in *H. melpomene* and *H. erato*, as it stood at the start of this project, can be found in Figure 2.1.

Several populations and species of *Heliconius* have shared their wing pattern alleles by introgression. The species *H. heurippa*, which has a red and yellow forewing band, gained the red portion of the pattern via introgression from red-banded *H. melpomene* in a process known as hybrid trait speciation (Mavárez et al 2006). Multiple studies have described introgression events between *H. melpomene* and *H. timareta* in the Andes, as well as with the more distantly related *H. besckei* and *H. melpomene* in southern Brazil (Pardo-Diaz et al., 2012, Zhang et al., 2016b, Salazar et al., 2010). This sharing of pattern modules between populations leaves signals of introgression in the genome. This was used by Wallbank et al to identify the *dennis* and *ray* modules. The method of topology weighting by iterative sampling of sub-trees (*Twisst*) developed by Martin and van Belleghem (2017) can now be used to scan for topologies that support introgression at loci of interest.

It has been widely hypothesized that developmental evolution primarily occurs through regulatory change, involving novel functions for conserved regulatory and signaling molecules. The role of *optix* is a prime example of this, as this gene is involved in the rapid radiation of *Heliconius* wing patterns while also having a wide range of other functions. In butterflies, these include a more broadly conserved role in specifying a set of scales in the region of wing overlap, but also expression domains in the optic lobe and medulla of the pupal brain (Martin et al., 2014). Similarly, *WntA* is involved in wing patterning in multiple Lepidoptera (Martin et al 2014, Gallant et al 2014), and is also a constituent of the Wnt signaling pathway, which is vital for development, and the genes *cortex*, *domeless* and *washout* have been linked to segmentation defects in embryonic development (Saenko et al 2012).

Aims of this study

No regulatory sequence associated with the *band* element (Table 1.1) has yet been described, and at the time of this study, only a broad associated region had been described in *H. erato*, with no specific associations with individual pattern elements. Subsequent work published by Van Belleghem et al. somewhat replicated the work described here, and I will compare the results from the two studies in the discussion of this chapter. To investigate the functional sequence changes that lead to wing pattern evolution, we must identify the minimal sequence required to recapitulate patterned gene expression. I aim to describe new regulatory modules and narrow the previously described ones by using additional samples for association and introgression analyses. This will assist in the investigation of the functional basis of pattern variation, and will also allow a multispecies comparison of homology and convergent evolution in regulatory sequences. Armed with a full complement of mimicry modules for both species, I can perform a multi-species comparison of homology and convergence of these mimicry modules between the two species.

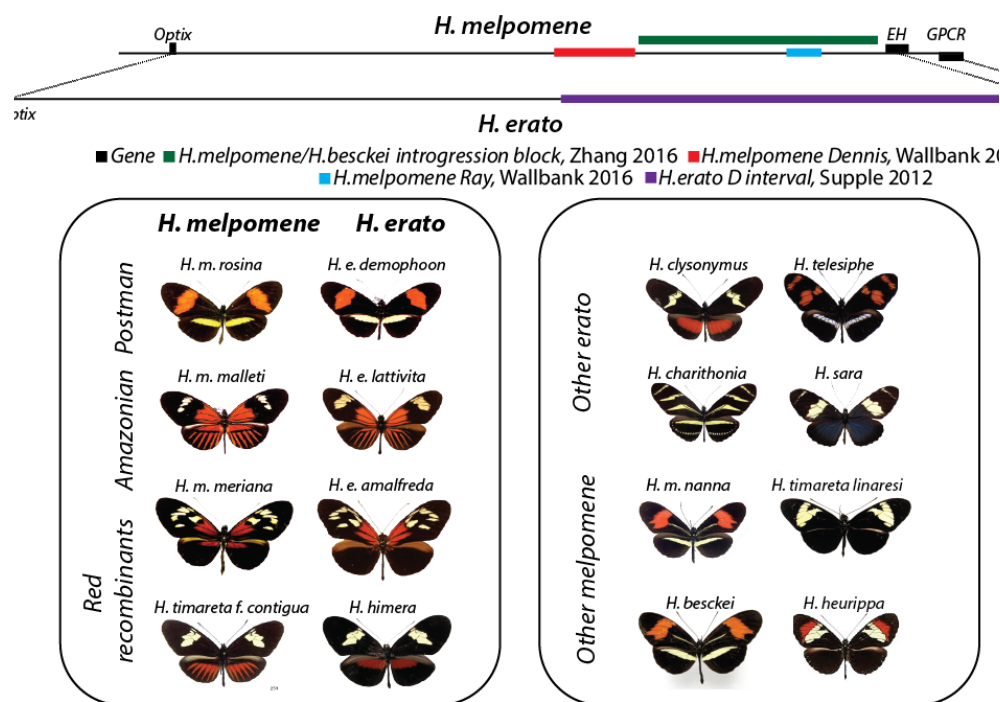


Figure 2.1: Summary of published data on the *optix* regulatory region in *H. melpomene* and *H. erato*.

TOP: Cartoon of *optix* locus in *H. melpomene*. Dotted lines indicate homologous genes. Previously published windows of pattern association are indicated by coloured bars: in purple, the 65 kb window of association in *H. erato* identified by Supple et al (2013); in red, the module associated with *H. melpomene* hindwing rays, and in blue, the module associated with forewing proximal red (called the dennis patch), as described in Wallbank et al (2016). In green, a window showing signals of introgression between *H. m. nanna* (southern Brazil) and the local co-mimic *H. besckei* (Zhang et al., 2016).

BOTTOM LEFT: The most widespread pattern forms of *H. melpomene* and *H. erato*, the Postman form (Red forewing band, yellow hindwing band), and the Amazonian form, (hindwing rays and the dennis patch, with yellow forewing band). Recombinant morphs of these red patterns can be found in both species groups. The Surinamese *H. m. meriana* and *H. e. amalfreda* both have the dennis patch but no hindwing rays, while *H. timareta timareta f. contigua* and *H. himera* both have red hindwing patterns but no dennis patch.

BOTTOM RIGHT: other pattern forms and species from the *H. erato* and *H. melpomene* groups.

II: METHODS

Sampling and sequencing, *H. erato* clade

I examined sequence variation of 19 taxa (n=69) across the broader *Heliconius erato/sara/sapho* clade, representing both convergent and divergent phenotypes (Figure 1.2). This included previously published data for 12 taxa (number of individuals (n)=58) (Supple et al., 2013), and Megan Supple collected and sequenced 7 new taxa (n=11) (Supplementary Table 1) for this study. The overall dataset consisted of multiple races of *H. erato*, representing the postman phenotype (6 races, n=32) and the dennis-ray phenotype (4 races, n=18). This included related species within the clade with the postman phenotype (1 species, n=2) and the dennis-ray phenotype (1 species, n=1) as well as *H. e. amalfreda* (n=5) and *H. himera* (n=5). Additional species were sampled with patterns similar *H. himera* (3 races, n=4) and species within the clade with no red colour elements (2 species, n=2). Samples were sequenced to at least 30x coverage using whole genome sequencing of 100-bp paired end sequencing reads on an Illumina HiSeq platform and single nucleotide polymorphisms were genotyped as described in Supple et al. (2013).

Improving quality of *H. erato* reference sequence

At the time of execution of this project, a high-quality reference genome for *H. erato* was not available. A BAC-walk incorporating the *optix* locus had previously been sequenced. However, the substantial amount of variation between races of *H. erato*, especially indels and transposable elements, results in substantial missing data when aligning sequencing reads to the *H. erato* reference sequence, which was generated with individuals from Panama. To improve sequence coverage within the putative enhancer regions, I used race-specific reference sequences that had been generated using long-range PCR by Carlos Arias in Panama. The race-specific references were aligned to each other, as well as the *H. e. demophoon* genomic reference for that region.

Identifying modular enhancers in *H. erato*

Megan Supple performed multiple population genomic comparisons to localize functionally important regions within the 65-kb block of strong divergence between Postman and Amazonian *H. erato* races (figure 2.1), purple), examining genotype by

phenotype association of sequences derived from alignment to a partial reference sequence as described in Supple et al. (2013). Briefly, this involved aligning sequencing reads to a partial genomic reference sequence using BWA alignment software. This included a calculation of per-position genotype-by-phenotype association using a two-tailed Fisher's exact test based on allele counts and identified SNPs showing perfect genotype by phenotype association. Supple filtered out positions if less than 75% of individuals were genotyped for each phenotype. Together we identified putative regulatory modules by looking for regions of high genotype by phenotype association, including the presence of perfectly associated SNPs. To ensure we kept potentially important flanking sequences, these regions were extended on either side to include the next called SNP that was identified as fixed in the initial analysis of *H. erato* hybrid zones.

Rays

Supple identified the functional region modulating the presence of hindwing rays by examining association between three races of *H. erato* from neighboring populations: *H. e. hydara* (postman), *H. e. erato* (rayed), and *H. e. amalfreda* (recombinant). *H. e. amalfreda* has a yellow forewing band and red dennis patch that is characteristic of the rayed phenotype, but similar to the postman phenotype it lacks the hindwing rays. We examined genotype by phenotype association of rayed samples (n=6) versus non-rayed samples (n=12). Additionally, to identify a narrower, high priority region within this area, we examined genotype by phenotype association across the broader *erato/sara/sapho* clade ($n_{\text{ray}}=19$, $n_{\text{noHWred}}=41$). Long-range PCR amplicons were used to *de novo* assemble and align sequences from each individual, allowing the identification of recombination breakpoints.

Dennis

Similarly, we identified the functional region modulating the presence or absence of the dennis element by examining association across the *erato* clade. In addition to the classic postman (no dennis patch) and rayed (dennis patch) in *H. erato*, this extended sampling included 3 species, including *H. himera*, that have a yellow forewing band, a red hindwing bar, but lack the dennis patch. We examined genotype by phenotype association between dennis samples (n=23) and non-dennis samples (n=42). We did an additional association analysis across the broader *erato/sara/sapho* clade

($n_{\text{dennis}}=24$, $n_{\text{no-dennis}}=45$). Long-range PCR amplicons were used to *de novo* assemble and align sequences from each individual, allowing the identification of recombination breakpoints.

Band

Finally, I localized the functional region modulating the color of the forewing band by extending our analysis across the *erato/sara/sapho* clade. This extended sampling includes additional species that have a yellow forewing band and no red color elements. The yellow forewing band appears to be the ancestral phenotype and it is shared with the otherwise derived rayed phenotype. We examined genotype by phenotype association between the yellow forewing band ($n=35$) and the red forewing band ($n=34$). Long-range PCR amplicons were used to *de novo* assemble and align sequences from each individual, allowing the identification of recombination breakpoints.

Phylogenetic analysis in H. erato

Supple used phylogenetic analyses to examine the history of *H. e. amalfreda* and *H. himera* relative to the larger *H. erato* radiation. First, she examined where *H. e. amalfreda* and *H. himera* fell within the *H. erato* radiation across the 65 kb regulatory region by generating non-overlapping neighbor joining trees for 5 kb windows from all 19 taxa ($n=69$) using PAUP* (Swofford, 2003), then used a reduced dataset to test the log likelihood of the data under alternative trees. Each tree had four *H. erato* races (*hydara*, *erato*, *favorinus*, *emma*) plus the race or species of interest. The five taxa were assumed to be monophyletic and all samples within the five race/species were unresolved relative to each other. For each comparison, the two trees only differed in their placement of the taxon of interest. For *H. e. amalfreda*, we tested whether it clustered with the rayed or the non-rayed samples. For *H. himera*, we tested whether it clustered with the dennis or the non-dennis samples. We examined sliding windows across the region, with 5-kb windows and 1-kb slide and determined the negative log likelihood of the data under each of the two hypothesis trees using PAUP* LScores. Using reference sequences generated from long range PCR, I assembled each *H. erato* and constructed phylogenetic trees in sliding windows to confirm recombination breakpoints between informative races at pattern modules.

Sampling and sequencing, *H. melpomene* clade

I utilized the data set previously generated by Wallbank et al (2016), consisting of 43 taxa (n=140) of convergent and divergent pattern forms of *H. melpomene*, *H. cydno*, *H. timareta*, *H. numata*, and numerous other taxa. This included whole genome resequenced individuals, as well as individuals that were sequenced in a selective sequencing protocol (SureSelect) by Nadeau et al (2012), only sequencing the *optix* locus, *cortex* locus, and two unlinked loci not associated with any known function. In addition, I analysed *H. besckei* (n=4) from Zhang et al (2016), as well as two additional *H. heurippa* individuals, bringing *H. heurippa* to n=4.

Identifying band module enhancer in *H. melpomene*

Genotype-by-phenotype analysis

I performed genotype-by-phenotype analysis on *H. melpomene*-clade individuals with and without the forewing red band in a genome wide association analysis using GenABEL. Each individual (listed in table S2.2) was aligned to the *H. melpomene* genome v2.0 with BWA, default parameters, and filtered for scaffold Hmel218003 (which contains HE670865 from v1.1, plus a number of other v1.1 scaffolds). Whole genome analysis could not be performed as some critical samples (*H. timareta linaresi*) were sequenced under a selective sequencing protocol. Variant loci were called using GATK v3.4-46. Probability scores for association were generated using the emp.qscore function, and plotted as $-\log P$. Sites with scores of $-\log P > 4$ or higher were considered significant. Windows including the identified sites were *de novo* assembled and aligned with MAFFT into a multiple alignment of all taxa. Maximum likelihood trees were generated using PhyML v2.2.0.

*Topology weighting using *Twisst**

Scaffold Hmel218003 was partitioned into sliding windows, each containing 50 SNPs. For each window, a phylogenetic tree of all individuals was constructed using PhyML v2.2.0 with default parameters. Each individual was assigned to a taxon, and sets of 5 taxa were selected for the calculation of topology weightings. Unrooted trees with 5 taxa have 15 possible topologies (here numbered 0-14, (supplementary figures S2-1 and S2-2). Topology weightings were calculated for each window, as described in Martin and Van Belleghem (2017). Briefly, *Twisst* calculates the relative contribution of each possible topology to a phylogenetic tree by iteratively

subsampling it, and summing the occurrence of each possible topology (shown in supplementary figures 2.1 and 2.2). Multiple unrooted topologies may support one given hypothesis of relationship; for example topologies 9 and 11 are both consistent with the species-level phylogenetic arrangement described by Kozak et al (2015), and therefore are both consistent with no introgression. On the other hand, topologies 3, 4 and 5 are all consistent with introgression between *H. heurippa* and red-banded *H. melpomene* to the exclusion of *H. t. linaresi* and *H. cydno*, and therefore are all consistent with introgression. These correlated topologies were summed at all positions and plotted by position along Hmel218003. The Species topologies were plotted as negative values (in grey), and Introgression topologies were plotted as positive values (in colour).

Analysis of conservation between H. erato, H. melpomene, and other Lepidoptera

The scaffolds containing *Optix*, *WntA* and *cortex* from *H. melpomene* and *H. erato* were aligned with MEGABLAST, with the alignment visualized using the Artemis Comparison Tool. These scaffolds were used to search the genomes of other Lepidoptera stored at LepBase using BLASTn – (*Bicyclus anynana*, *Danaus plexippus*, *Pieris napi*, *Plodia interpunctella* and *Bombyx mori*). I selected the genomic scaffolds in each species that contained the protein-coding gene, as well as co-linear BLASTn hits to non-coding sequence: this was assisted by the ‘Gene Tree’ homology function on LepBase. All seven species were then aligned using mVISTA, and a plot of conservation was generated. All sequences with at least 75% conservation and at least 70 bp in length were highlighted.

III: RESULTS

Identification of pattern-associated modules in H. erato

Using genotype-by-phenotype association mapping across the region previously described by Supple in *H. erato* (Supple et al., 2013), we identified three genomic regions associated with *dennis*, *ray* and *band* respectively.

Ray

To identify the *H. erato ray* module, we analyzed five races (n=18) that differ in wing pattern across the continuous range of this species (figure 2.2A). These were *H. e. emma* and *H. e. erato* that have the hindwing ray phenotype, and *H. e. hydara*, *H. e. favorinus* and *H. e. amalfreda* that lack hindwing rays. The populations sampled are from across the range of the species – Peru (*emma* and *favorinus*) and the Guiana Shield (*hydara*, *amalfreda* and *erato*) largely controlling for any geographic structure. Across these samples we identified 12 SNPs that are perfectly associated with the presence or absence of hindwing rays. Of these, 11 are within a 12 kb window (Figure 2.2a) and the remaining site is adjacent to *optix*. Additionally, two of these SNPs remain perfectly associated when the analysis is extended to include all taxa with red hindwing patterns across the broader *H. erato* clade. The SNP associations were complemented by an analysis of tree likelihoods where we searched specifically for regions where *H. e. amalfreda* switches phylogenetic position. This analysis highlighted a 16 kb region which completely overlapped the 12 kb window identified by SNP associations (Figure 2a).

I next used *de novo* assembly and manual alignment of the focal region to identify precise breakpoints associated with the hindwing rays phenotype. This analysis included long-range PCR products that were sequenced and individually assembled to provide high quality reference sequence for specific wing pattern forms. One form, the race *H. e. amalfreda*, was particularly informative as it possesses the *dennis* patch but lacks hindwing rays (figure 1). This form shares the haplotype of Amazonian races across most of the genomic region, but switches to a Postman haplotype in the region already identified as showing ray-associated SNPs. These recombination breakpoints specify a candidate *ray* enhancer region of 12 kb. Phylogenetic trees of

this region supported the result and demonstrated that *H. e. amalfreda* forms a clade with the red-banded Postman races only in this region. This implies that the *H. e. amalfreda* phenotype is a result of recombination between Postman and Amazonian alleles, similar to that reported previously in the co-mimic *H. m. meriana* (Wallbank et al., 2016).

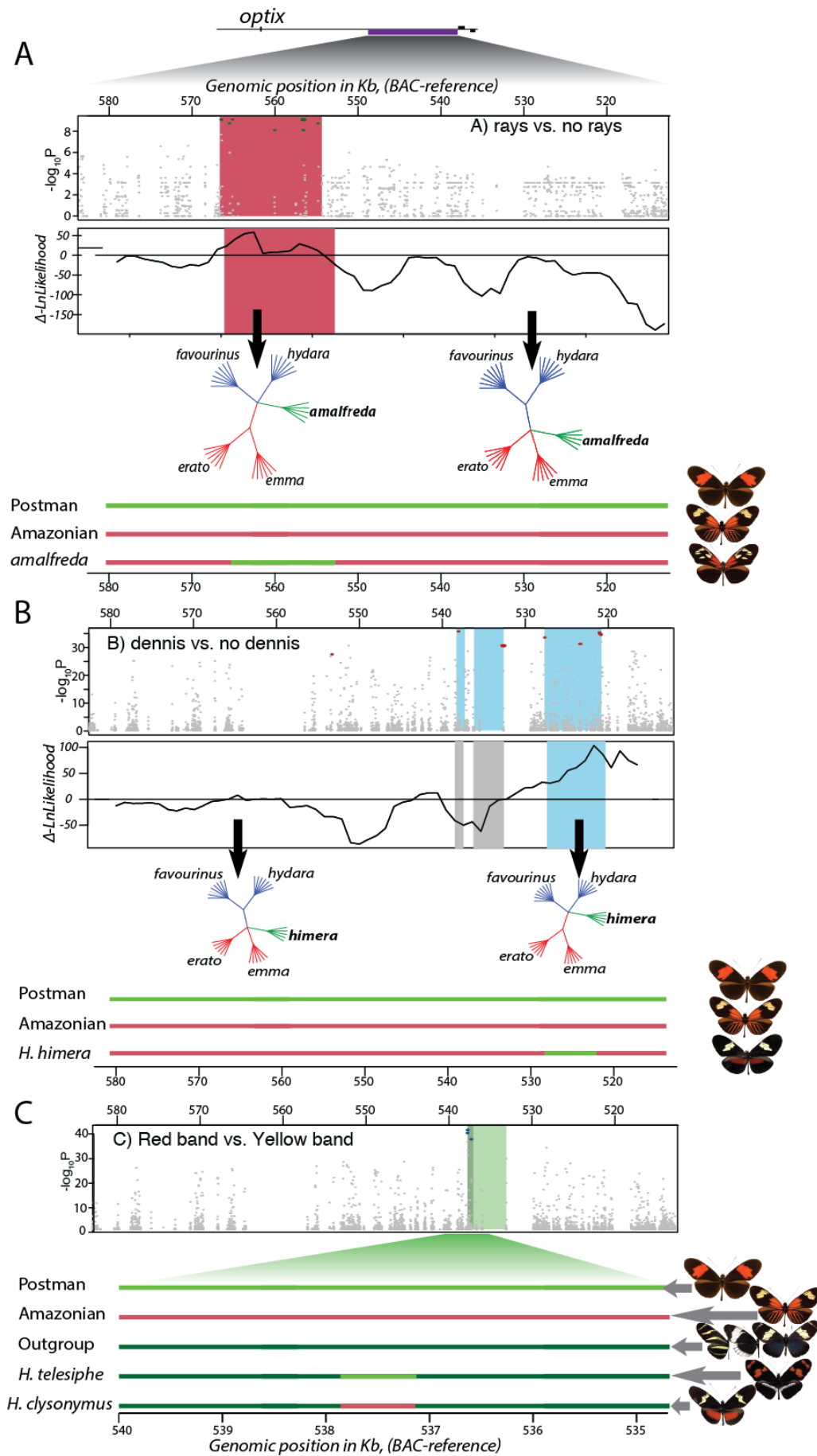


Figure 2.2: Genotype-by-phenotype analyses of wing patterns in *H. erato*

Genomic scale indicators use the *H. erato* BAC-walk (and so run right-to-left).

A, ray element. One SNP cluster with association to the *ray* element was identified (shaded red), with fixed SNPs indicated in green. The neighbor joining 5-kb trees show *H. e. amalfreda* clustering with the rayed races, except for two adjoining 5-kb windows where *H. e. amalfreda* switches to cluster with the postman races. The likelihood analysis favors *H. e. amalfreda* clustering with the rayed forms across most of the region, except a 16-kb region, where the alternative tree, with *H. e. amalfreda* clustering with the Postman forms, is favoured.

B, dennis element. Genotype-by-phenotype analysis identified three clusters of associated SNPs (coloured red, in grey and blue windows). The maximum likelihood scores favour the clustering of *H. himera* with the Amazonian allele across most of the region, except for a 7 kb window which matches one of the three SNP clusters.

C; band element. Genotype by phenotype analysis of red vs yellow band individuals identified a cluster of just three SNPs in a 34bp window, indicated by the dark green shaded box. The window is extended to the next SNP identified in the hybrid zone analysis by Supple et al (2013) to account for a large window of poor mapping (lighter green).

Dennis

For the *H. erato dennis* region, I carried out a similar set of analyses but also included the closely related species *H. himera* which has a ray-like hindwing red bar but no corresponding forewing dennis patch, similar to the situation in *H. melpomene*. Genotype-by-phenotype analysis of 15 taxa including *H. erato* and *H. himera* identified 10 SNPs perfectly associated with the presence or absence of the forewing dennis patch. Of these, 7 are located in a core 7 kb region; the other 3 are within the broader 65 kb regulatory region. Based on these SNPs we highlighted initial candidate regions (figure 2.2b, grey and blue boxes). We then used a maximum likelihood analysis with a subset of just five taxa to identify a region of approximately 7 kb in which *H. himera* groups with the postman phenotypes that lack the dennis patch (Figure 2B). The two additional regions identified from SNP associations were ruled out as they showed no phylogenetic signal and in both cases had been identified based on a single perfectly associated SNP. Finally, I again used *de novo* assembly of shotgun sequences, combined with long-range PCR sequencing to similarly construct a high-quality alignment of the focal region and delimit recombination breakpoints for the *H. himera* allele. This confirmed the region of 7 kb in which *H. himera* has a Postman allele in a genetic background otherwise similar to the Amazonian forms. Phylogenetic reconstruction of all taxa in the analysis confirmed this pattern.

Band

Next we focused on the *band* region. Using genotype-by-phenotype analysis on 19 taxa across the *erato/sara/sapho* clade comparing red versus yellow forewing bands, we identified 4 SNPs in a 5 kb region that were perfectly associated with forewing band colour (figure 2C). Three of these SNPs clustered in a 34 bp window. *De novo* assembly and alignment confirmed this pattern and in particular identified the red-banded outgroup species, *H. telesiphe* as especially informative. Despite the fact that this species diverged ~5.5 Ma ago from *H. erato* (Kozak et al 2015), it shows an erato band-like haplotype across a region of 11 kb overlapping with the identified associated SNPs. In surrounding genomic regions *H. telesiphe* is more similar to its sister species *H. clysonymus* and *H. hortense*, which have yellow forewing bands. This therefore represents a novel recombination event across a considerable genetic distance, and likely indicates the occurrence of a novel introgression event (figure 2.3).

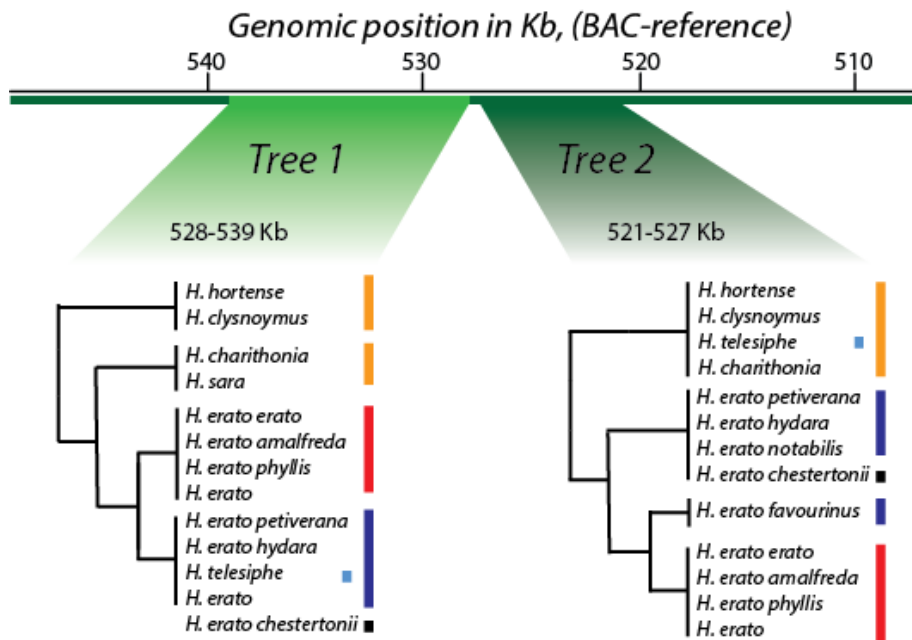


Figure 2.3: Maximum likelihood trees of the *H. erato* band element

Maximum likelihood trees show *H. telesiphe* groups with Postman erato within the 11 kb band window, whereas in an adjacent window of 8 kb, *H. telesiphe* groups with its sister species *H. clysonymus* and *H. hortense*, as it does in the phylogenetic reconstruction by Kozak et al (2015). Notably, *H. erato chestertonii* also groups with the Postman forms of *H. erato* in this module despite having no red mimetic pattern elements.

Identification of the band module in H. melpomene

In the co-mimic species *H. melpomene*, previous work has already identified *dennis* and *ray* modules (Wallbank et al 2016). I here focus on identification of a band module to complete the set of three mimetic patterning modules in each of the two clades. I first compared red-banded individuals of *H. melpomene*, *H. timareta*, *H. heurippa* and *H. besckei* to yellow-banded individuals of *H. melpomene*, *H. cydno* and *H. timareta* in a GWA analysis of the chromosome containing *optix*. Genotype-by-phenotype association identified 20 associated SNPs in two clusters which were shared by all red banded individuals in all tested species ($p < 0.0001$) (Figure 2.4), (SNPs highlighted in red).

H. heurippa has a compound red/yellow forewing band which it has gained by adaptive introgression with *H. melpomene*, and *H. besckei* shares the Postman phenotype with *H. melpomene* despite being a member of the divergent silvaniform clade, other members of which do not share the Postman pattern form. Both species are hypothesized to have gained the red through adaptive introgression (Salazar et al., 2005, Mavárez et al., 2006, Zhang et al., 2016a). In order to investigate this signal of between-species introgression, I used the topology weighting program Twisst to screen the genomic region.

First, *H. heurippa* and its yellow-banded sister population *H. timareta linarensis* were compared to *H. melpomene* and *H. cydno* (Figure 2.5). Across most of the chromosome, the topologies conformed to the null expectation of non-introgression (grey, Figure S2.1). In the region around SNP cluster 2, though not around SNP cluster 1, the weighting for these non-introgression topologies is reduced to near-zero and the topologies which support introgression become the most strongly supported reaching 1.

Next, red-banded *H. melpomene*, including *H. m. nanna* from southern Brazil, were compared to three silvaniform populations, the tiger-striped *H. pardalinus* and *H. sergestus*, and the red-banded *H. besckei* (Figure 2.5). As above, non-introgression topologies had high support across most of the chromosome, except around SNP cluster 1 and 2 where the introgression topologies were more strongly supported. The region of SNP cluster 1 is around 12 kb in length, and the region of SNP cluster 2 is

20 kb.

Together, these findings suggest that there are two distinct *band* modules in the *H. melpomene* clade; one of these was shared between *H. melpomene* and *H. heurippa* and was sufficient to cause the gain of a red band, and an additional module, closer to *optix*, which was shared together with the *H. heurippa* module between *H. melpomene* and Silvaniform species. This data has been corroborated by further studies of introgression between *H. melpomene* and other Silvaniform species carried out by Morris et al, who were able to extend this method to detect additional signals of introgression at the other major colour pattern loci *cortex* and *WntA* (Figure 2.6).

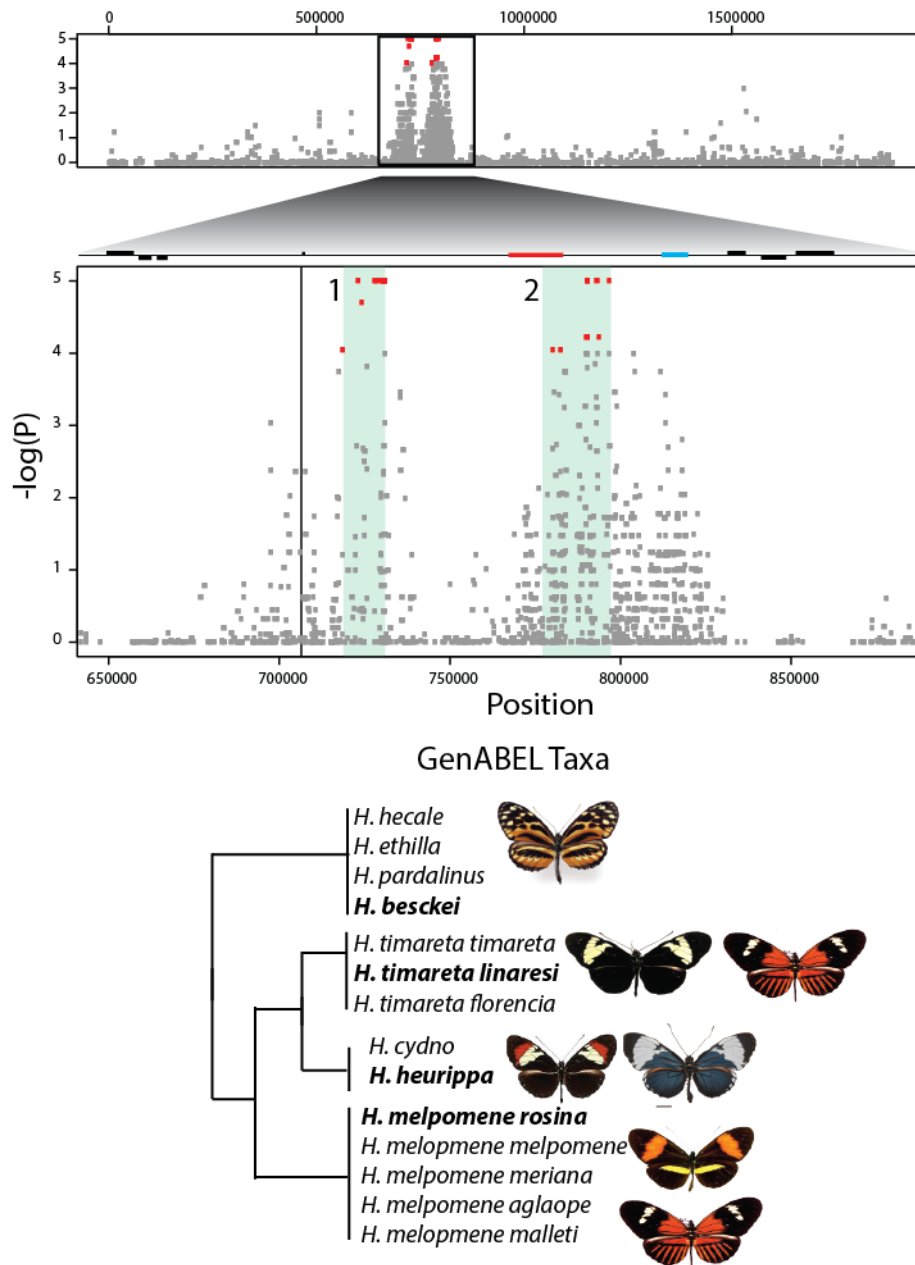


Figure 2.4: Upper panel shows permuted probability of association across the whole Hmel218003 scaffold, indicating the presence of two distinct peaks of SNPs. (Note that SNPs with a $-\log(P)$ score of 0 are not included in the figure). Lower panel shows the two peaks, labeled 1 and 2, with the annotation of *optix* and regulatory elements from Wallbank et al (2016), with green shading on the SNP peaks, and SNPs with $-\log(P) > 0.0001$ coloured red. The phylogeny indicates all taxa included in this analysis, also listed in supplementary tables.

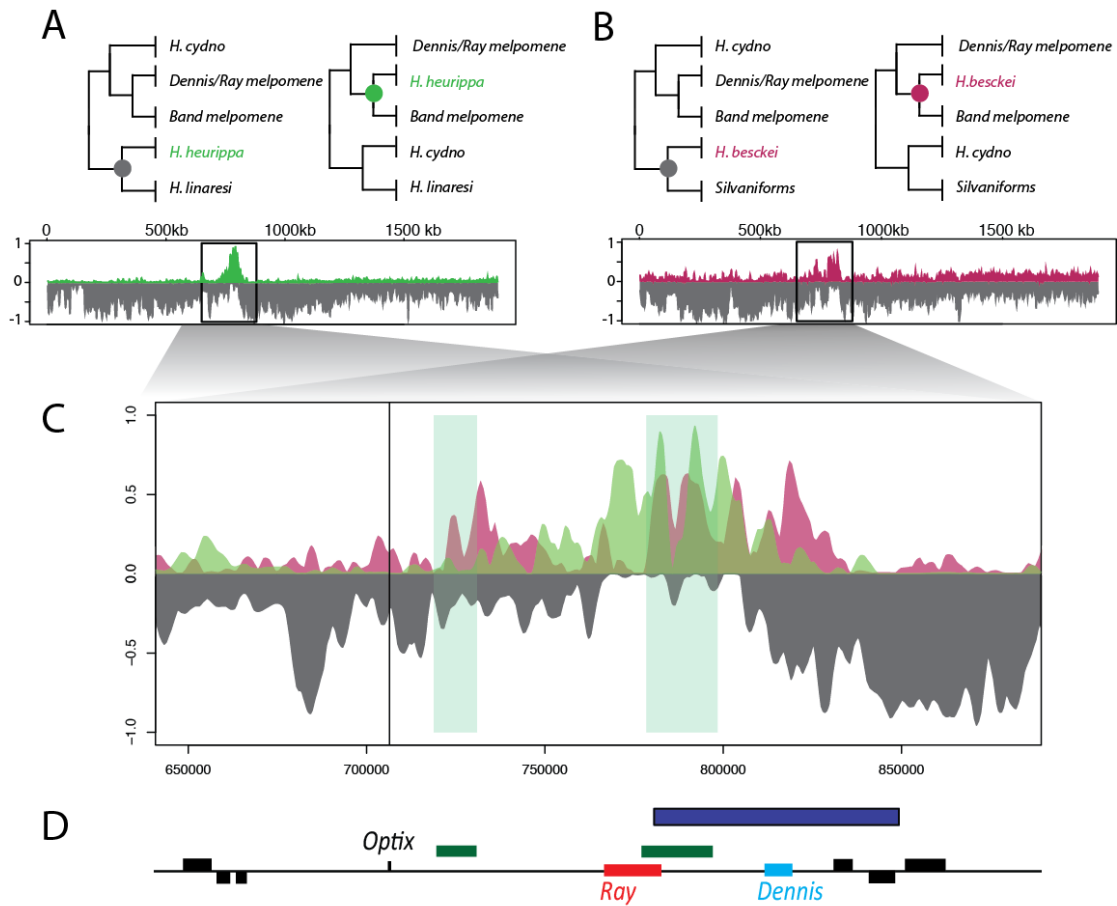


Figure 2.5: A and B; To the left, the ‘species topology’ and to the right, the ‘introgression topology’. In A, note the movement of *H. heurippa* from a monophyletic clade with *H. timareta linaresi* to a monophyletic clade with *H. melpomene*. In B, note the movement of *H. besckei* from a monophyletic clade with silvaniforms to a monophyletic clade with *H. melpomene*. Topology weightings; positive values indicate support for introgression topologies, while negative grey values indicate support for the species topology. C, overlaid topology weightings for the two comparisons, with SNP windows 1 and 2 from Figure 2.4 indicated by dark green shading. D indicates the annotation of this locus, with the elements identified by Wallbank (2015) indicated in blue and red, and the region identified by Zhang et al (2016) as introgressed between *H. melpomene* and *H. besckei* indicated in purple. Note that a value of 1 or -1 indicates that a particular topology shows reciprocal monophyly among the sampled individuals.

Non-coding sequence conservation around wing patterning genes

The identification of the *band* module in *H. melpomene*, as well as the *band*, *ray* and *dennis* modules in *H. erato*, means that the complement of all red pattern modules has now been identified in both species. Following this project, Van Belleghem et al (2017) were able to use *Twisst* to recapitulate this identification of pattern modules in *H. erato*, using a dataset with additional samples and with the highest-quality reference genome assembly of any lepidopteran, with very few unassembled gaps. Independently, they found multiple modules associated with the hindwing bar (near *cortex*) and with forewing band shape (near *WntA*), and were also able to identify multiple modules at the *optix* locus, including additional elements that were not identified in this analysis. In addition to this, Morris et al (in submission) have used additional introgression events between silvaniform species and *H. melpomene* to identify wing pattern modules at *cortex* and *WntA* using *Twisst*. This allows for a direct test of whether these pairs of mimetic wing pattern modules have evolved within the same regulatory modules, or via some non-homologous mechanism. All currently-described modules are listed in the table below.

Locus	<i>H. erato</i>		<i>H. melpomene</i>	
	Module	Role	Module	Role
<i>Optix</i>	D	shape of dennis element	<i>dennis</i>	shape of dennis element
	R	shape of ray element	<i>ray</i>	shape of ray element
	Y	defines colour of band element	<i>band</i>	shape of band element
<i>WntA</i>	Split (St)	shapes central part of forewing band		
	Shortened (Sd)	patterns proximal part of forewing band		
	Yellow line (Ly)	patterns distal part of forewing band		
			module 1	??
			module 2	??
<i>Cortex</i>	Cr ^(west)		Yb (yellow bar)	patterns hindwing yellow bar
	Cr ^(east)			
			Melanic tip	patterns the melanic tip of the forewing
			Yellow band	patterns forewing yellow band
<i>Vvl</i>	Ro	melanic tip	(QTL)	

Table 2.1: List of elements that have now been identified in each species at each locus.

While homologous sequence has previously been identified in the *optix* regulatory region (Supple et al, 2013), the positional homology and orientation of this homologous sequence was not shown. In order to observe the arrangement of homologous sequence, and to see if pattern modules were contained within collinear sequence, I first aligned the genomic region from the *H. melpomene* and *H. erato* *optix*, using MEGABLAST, and visualized with the Artemis Comparison Tool (Figure 2.6). This demonstrated complete synteny between the two species across this region. There is no evidence for major rearrangements apart from one 20 kb duplication in *H. erato* relative to *H. melpomene*. Additional modules associated with pattern elements were identified by Jake Morris at the *cortex* and *WntA* loci. As such, I also aligned these genomic loci between *H. melpomene* and *H. erato*. Again, there is no evidence for major rearrangements at these loci, though an approximately 5 kb tandem repeat is present in both species at the *WntA* locus just 3' of the gene.

At the *optix* locus, both the *ray* and *dennis* modules were localised by BLAST to positionally homologous locations in *H. melpomene* and *H. erato*. The *Ray* elements contain a substantial amount of conserved sequence between the two species, as well as a short region which is deeply conserved between other Lepidoptera, implying the presence of an ancestrally shared regulatory element within the *ray* element of both species. The *dennis* module contains homologous sequence in both species, with an overlap of 5034/7599 bp relative to *H. erato* (Figure 2.7). Note that no SNPs are shared between species.

The *H. melpomene* *band* modules identified herein are located at a substantial distance from the corresponding module identified by Van Belleghem et al.. While these modules have high sequence conservation between *H. melpomene* and *H. erato*, they have very low levels of conservation between the other Lepidopteran species. One of the *Band* modules is partly situated in a 35 kb region which is duplicated in *H. erato* relative to *H. melpomene*. The physical distance between the modules in the two species indicates that the evolution of the red band occurred by acquisition of regulatory changes at unrelated loci.

At the *WntA* locus, module 2 contains homologous sequence with the *H. erato* *Ly* element. This module contains two peaks of conservation with multiple other

lepidopteran species, which correspond to the two 3' coding exons of the *WntA* gene. Module 1 does not contain homologous sequence to any of the *H. erato* WntA modules, and is not enriched for sequence conservation with other species versus the surrounding region, indicating that it represents an independent and non-convergent locus. At *cortex*, the *H. melpomene* yellow bar module was contained within one of the two large yellow bar-linked regions identified in *H. erato* by van Belleghem et al. Notably, in both species this module includes the 3' portion of the gene *parn*, which codes for a polyadenylate-specific ribonuclease. The *H. melpomene* yellow band modules identified here contain the exons of the gene *cortex*.

In order to ascertain whether wing pattern elements tend to evolve in ancestrally conserved regulatory modules, I investigated patterns of sequence conservation across the *optix* regulatory region throughout the Lepidoptera. The *Heliconius optix* region was aligned by BLASTn with the Nymphalid butterflies *Bicyclus anynana* and *Danaus plexippus*, the Pierid butterfly *Pieris napi* and the moths *Bombyx mori* and *Plodia interpunctella* (downloaded from from LepBase). We detected 20 non-coding multispecies conserved sequences (MCS) shared between all species, and a further 22 conserved between *Heliconius* and at least one other Nymphalid butterfly. Clusters of MCS are present within the identified pattern modules, suggesting that these regions have ancestral functions in other species of butterfly and moth (figure 2.6).

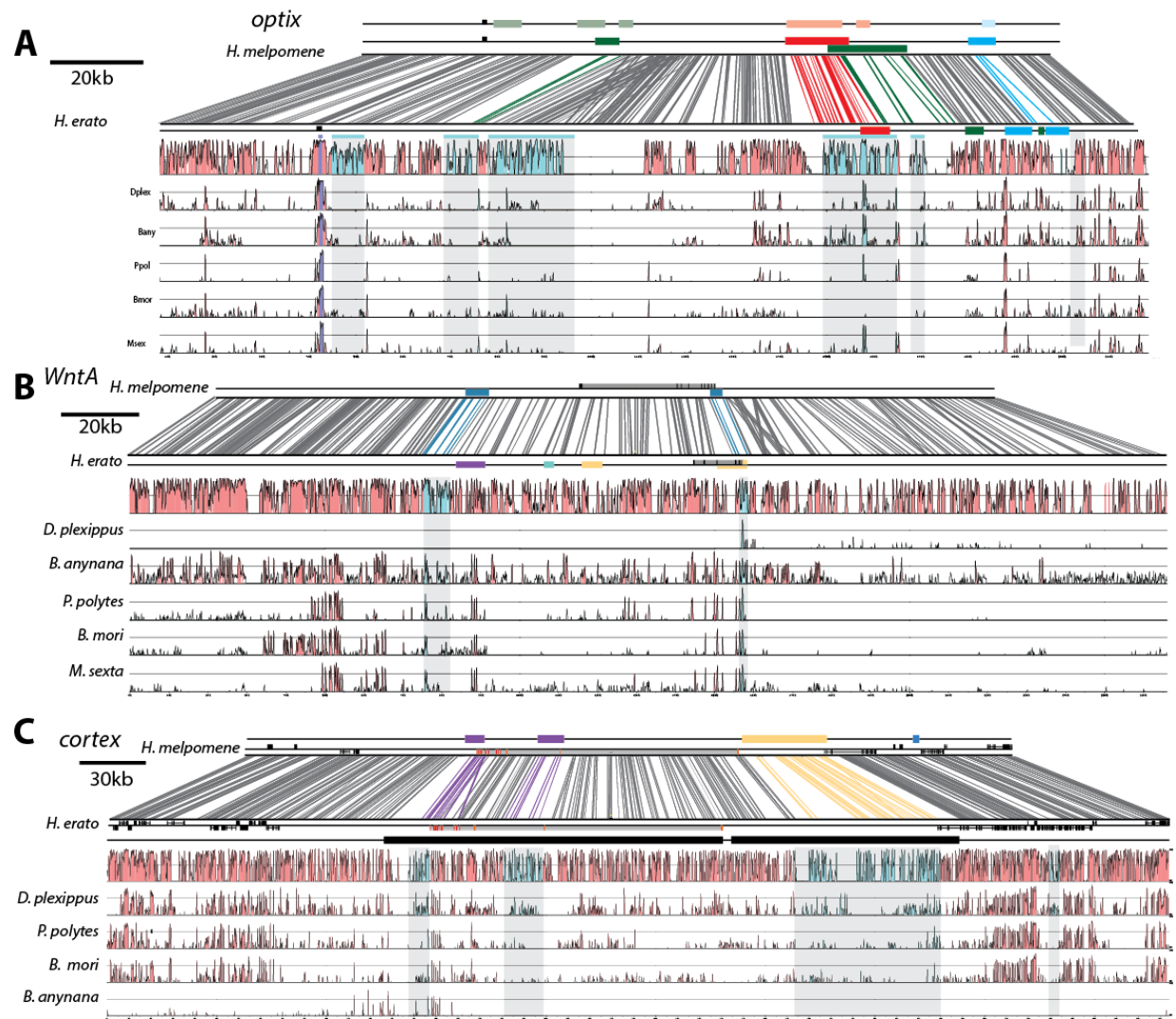


Figure 2.6: Multi-species conservation analysis of the *optix*, *cortex* and *WntA* loci

A shows *optix*, B shows *WntA* and C shows *cortex*. UPPER Artemis representation of MEGABLAST alignment of *H. melpomene* and *H. erato* *optix* regulatory region illustrating colinearity of all conserved elements (BLAST hits with similarity above 95% are shown).

LOWER mVISTA conservation plots for *H. melpomene*, *H. erato*, *B. anynana*, *D. plexippus*, *P. napi*, *Pl. interpunctella*, & *B. mori*. This indicates the presence of many multi-species conserved sequences, some of which co-localise with the identified regulatory modules (peaks with over 75% identity and longer than 50bp are coloured pink). The locations of these peaks are also indicated on the upper BLAST alignment, using coloured bars.

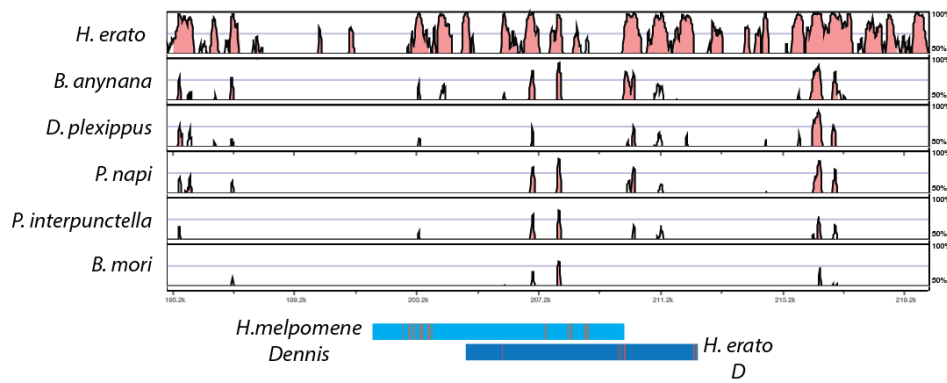


Figure 2.7: A closer look at the overlapping *dennis* modules. SNPs are indicated by red dashes, and show that there is no evidence for shared polymorphisms between the two species, nor for closely apposed polymorphisms. In the overlapping region, there are elements which are conserved between *Heliconius* and the other Lepidoptera. In *H. melpomene*, these elements contain two clusters of SNPs, but this is not the case in *H. erato*.

IV: DISCUSSION

It is clear that *optix* plays a critical role in the gene regulatory network that leads to the development and diversity of wing pattern. In order to understand how *optix* is integrated into the wing gene regulatory network in a way that creates expression domains for the band, ray and dennis pattern elements, it is necessary to identify regulatory sequences associated with this gene. The ultimate goal is to locate the minimal enhancer sequence which produces a given expression domain, and to determine the identity of the regulatory factors that bind there. In order to do this, sequences which are associated with pattern elements must first be identified, and in *Heliconius*, this can be achieved with association analyses and by identifying introgression events associated with particular patterns.

Here, I identified sequence associated with the red band element in *H. melpomene*, and with the *band*, *dennis* and *ray* elements in *H. erato*. I was also able to identify a novel candidate between-species introgression event, the first in the *H. erato* clade. While I could not functionally test these associated modules, I looked for sequence conservation within them as a proxy for functional sequence. This led to two main findings. Firstly, I identified short sequences throughout the *optix*, *cortex* and *WntA* regulatory regions which were highly conserved throughout the Lepidoptera, and which are therefore candidate functional sequences which may control expression of *optix*, as regulatory regions are sometimes conserved in sequence, co-linearity and function (Indjeian et al., 2016, Elgar and Vavouri, 2008). A number of these highly conserved elements sit within the pattern modules and may be associated with wing patterning function. Second, I have demonstrated that phenotypic convergence in butterfly wing patterns can be associated with similar sequence changes in different lineages, but also may occur via different mechanisms. The *dennis* module has evolved convergent function in homologous sequence in both *H. melpomene* and *H. erato*, while *band* has done so in non-homologous sequence. This suggests that *dennis* may be controlled by a ‘regulatory hotspot’. Likewise, the 3’ region of *WntA* is associated with a pattern module in both species, suggesting it is also a regulatory hotspot. This represents a substantial progression on the previous understanding of regulatory convergence in *Heliconius* (Supple et al., 2013).

Together, these findings suggest that two alternate modes of convergent evolution have occurred; in the case of *dennis*, convergent expression profiles of *optix* were generated through re-use of the same ancestrally conserved regulatory sequences, while in the cases of *band*, convergent expression profiles of *optix* were generated through modifications to either different ancestral regulatory modules or by generation of *de novo* modules.

It seems likely therefore that the forewing *dennis* patch, like lactase persistence in humans, has evolved by point mutation at the same pre-existing regulatory modules in both species. However, unlike the case of lactase persistence, no fixed polymorphisms are shared between the species indicating that the requisite changes to the *optix* expression profile were driven by a different set of mutations. *optix* has multiple expression domains which are unrelated to mimicry, including expression in the optic lobe, in the wing overlap region, and in the ventral hindwing dots, and these domains are ancestral within the genus (Martin et al., 2014). Additionally, the metazoan homolog of *optix*, *Six3*, is known to be involved in specification of apical neural fates throughout the eumetazoa and is a component of conserved Gene Regulatory Networks, implying the likely presence of multiple additional ancestrally conserved regulatory domains, all of which will require additional regulatory architecture (Steinmetz et al., 2010). It is possible that part of this ancestral set of regulatory modules has features, such as chromatin conformation or pre-existing GRN interactions in combination with the strong selective pressure to create a mimetic pattern of expression, that make it more likely for mutations that lead to the *dennis* expression profile to occur there in a form of developmental bias. I would suggest that such regulatory hotspots may occur in regions that contain pre-adapted sequence changes, which potentiate sequence to be more likely to acquire a new function, as described in the acquisition of citrate metabolism in experimental evolution of *E. coli* experiment (Lenski et al 2014).

Unlike *dennis*, the *H. erato* and *H. melpomene band* module has no overlap in sequence between the two species. As there is no evidence for sequence transposition or duplication, this indicates that the same expression profiles have evolved by independent mechanisms. The identified modules contain sequence conservation between *H. melpomene* and *H. erato*, as well as multispecies conserved sequences.

This means that the sequence in which these modules has evolved was probably not naïve, non-functional sequence but rather contains some pre-existing, ancestrally shared functional sequences. It is possible that the structure or function of ancestral *cis*-regulatory modules, as inferred from multispecies conserved sequences, have created a bias towards gaining new regulatory interactions at such sites. CRISPR-cas9 mediated mutagenesis of *WntA* has indicated that a factor downstream of the Wnt pathway must regulate the expression of *optix* in the *H. erato* *band* pattern, though the identity of the protein-DNA interaction that causes this regulation has not been elucidated. Interestingly, in *H. melpomene*, the *band* pattern does not appear to be affected by *WntA* mutagenesis, meaning that it is likely that other factors are responsible for regulating *optix*. It is possible that the clear difference in the functional basis for regulation of *optix* in relation to the *band* element is reflected in the spatially separated *band* modules in the two species.

New regulatory modules could arise stochastically and neutrally in naïve non-functional sequence, or they could evolve by modification to pre-existing regulatory modules, which already have the structure and function necessary to act as a regulator. Modification of a pre-existing regulatory module would require fewer mutational steps as opposed to *de novo* generation of a regulatory module, increasing the probability that a convergent novel function will evolve at that site; this ‘pre-adaptation’ was observed in the evolution of citrate metabolism in populations of *E. coli* in the Long Term Evolution experiment (Blount et al., 2008). Though no SNPs at the overlapping *dennis* elements were found to be identical or even very closely apposed, it is possible that a broad pre-existing functional domain was present in the ancestor which has been convergently modified, though by different means, to create the same result.

There are two related hypotheses for how pre-adaptation might occur – either the presence of any ancestral regulatory sequence could increase the likelihood of the gain of any new function at that locus, or alternatively, specific pre-adaptations increase the likelihood of specific convergent gains. There are few well-studied examples of constructive convergence in regulatory sequence, as few *cis*-regulatory mutations that pertain to convergent phenotypes have been identified, and so it is not yet possible to infer any general trend of whether one of these scenarios is more

probable than the other. A full picture of the regulatory landscape at *optix* will require some enhancer bashing or mutagenesis experiments, and with recent technological advancements, this may become possible in the near future.

Comparison with Van Belleghem et al modules for H. erato

The study by Van Belleghem et al (2017) described elements which overlap tightly with those described herein, but also identified two additional modules associated with *dennis* and *band* respectively. This study utilized a high quality reference genome for *H. erato* which was not available during the original analysis. Instead, I used a BAC reference sequence which contained a much higher percentage of ambiguous sequence and gaps than the reference genome. This meant that when illumina sequenced individuals were aligned to the reference, there were large regions in which no SNPs could be called, and this led to pattern-associated sequence not being identified.

Co-linearity of pattern modules

In both *H. melpomene* and *H. erato*, two discontinuous band modules have been identified. There are a number of scenarios that explain this. First, it is possible that the red band has evolved twice in the *H. melpomene* lineage, meaning there are two distinct modules – one shared with the silvaniforms, and a different one shared between *H. melpomene* and *H. heurippa*. This would reflect the discovery that the hindwing yellow bar *Cr* has evolved twice in *H. erato*, once east of the Andes and once West of the Andes (Maroja et al., 2012, Van Belleghem et al., 2017). Second, it is possible that the band module is not a discrete module; under some scenarios, it is possible for selection to act against modularity and clustering in regulatory sequence (Wagner et al., 2007). Thirdly, it is possible that there are two separate red band modules which are redundant to each other and each independently sufficient, or that one is necessary for the function of the other but that only one needs to be passed between lineages by introgression for the new functional expression domain to be gained (i.e. a form of regulatory epistasis).

Evolutionary history of recombinant phenotypes

We present the first evidence for enhancer shuffling between lineages with divergent wing patterns in the *H. erato* group, mirroring that described by Wallbank et al in *H.*

melpomene (2016). The race *H. e. amalfreda*, which has the Amazonian-like Dennis patch but no hindwing rays, has lost the *ray* module through introgression with Postman forms, in effect “shuffling out” the *ray* element. Similarly, *H. himera*, which has a red hindwing bar homologous to the Amazonian hindwing rays, has lost the dennis patch through introgression with Postman forms, this time “shuffling out” the *dennis* module. These recombination events create new alleles which are selectively advantageous in the context of the local mimicry assemblage. In this scenario, a “loss of function” of *optix* has occurred without the requirement for inactivation of a regulatory module by point mutation. Thus rapid diversification through recombination can act as the substrate for selection (Brown, 1983; Gilbert, 2002; Wallbank et al., 2015).

In addition to this within-species enhancer shuffling, we present evidence of between-species introgression of the *band* module in both *H. melpomene* and *H. erato* clades. Previous work has led to the inference that *H. heurippa* gained its red band from *H. melpomene* in a process of hybrid trait speciation, (Salazar et al., 2005, Mavarez et al., 2006). In addition, recent genomic evidence has been provided for introgression at the *optix* locus between *H. melpomene* and *H. besckei* in southern Brazil (Zhang et al., 2016). In both cases, the same module has passed between species through adaptive introgression. While the block of introgression from each species was non-orthologous, there was considerable overlap, and we can infer that the functional module is contained within this overlap.

Mirroring this, I provide evidence that *H. telesiphe* has gained its red band via introgression with *H. erato*. This is the first example of between-species adaptive introgression in the *erato* clade. However, it is also possible that the presence of a conserved red band allele between *H. erato* and *H. telesiphe* could be adequately explained by independent lineage sorting, or strong purifying selection maintaining this allele in a highly conserved way. One clear experimental test would be to sequence more races of *H. telesiphe* –currently only samples of *H. telesiphe sotericus*, from the Colombian Andes have been sequenced, which have a red forewing band, but the form *H. telesiphe cretacea*, from Peru, has a white forewing band. Sequences from both forms would allow further analyses, including *Twisst*, which would clarify whether this sharing of the red band was caused by an introgression event. Together,

these introgression events provide further supporting evidence for the role of introgression and recombination in the generation of diversity.

As well as providing evidence for introgression and recombination in pattern diversification, we also observed convergent red pattern element loss that cannot be explained by recombination. The race *H. erato chestertonii*, which is found in central Colombia and is unique amongst the *H. erato* races in lacking any mimetic red pattern elements (as it mimics *H. cydno gustavi* rather than any *H. melpomene*), groups with the Postman forms of *H. erato* in maximum likelihood trees, and also shares many fixed SNPs with Postman individuals across the *Optix* region. Similarly, the pattern form *H. timareta timareta f. timareta*, which also lacks red mimetic pattern elements, also groups with the Postman forms of *H. melpomene* and *H. timareta* across the *Optix* region. It is probable that both *H. e. chestertonii* and *H. t. timareta f. timareta* have lost the red band not through introgression and recombination, but through either point mutation or deletion.

Summary

By finding regulatory modules associated with wing patterns in these two species, we have begun the construction of a wing pattern Gene Regulatory network. In particular, I have shown that the convergent gains of the regulatory linkages that lead to pattern modification in each lineage can occur by repurposing the same sequence, or by using different sequences. Additionally, I have provided further evidence that introgression can generate evolutionary novelty without the need for mutation. The next crucial step in decoding the wing patterning GRN is to predict which proteins are binding in these regulatory modules to generate novel expression patterns.

References

- BLOUNT, Z. D., BORLAND, C. Z. & LENSKE, R. E. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A*, 105, 7899-906.
- BRAKEFIELD, P. M. 2006. Evo-devo and constraints on selection. *Trends Ecol Evol*, 21, 362-8.
- DINWIDDIE, A., NULL, R., PIZZANO, M., CHUONG, L., LEIGH KRUP, A., EE TAN, H. & PATEL, N. H. 2014. Dynamics of F-actin prefigure the structure of butterfly wing scales. *Dev Biol*, 392, 404-18.
- DOBLER, S., DALLA, S., WAGSCHAL, V. & AGRAWAL, A. A. 2012. Community-wide convergent evolution in insect

- adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proc Natl Acad Sci U S A*, 109, 13040-5.
- ELGAR, G. & VAVOURI, T. 2008. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet*, 24, 344-52.
- FRANKEL, N., WANG, S. & STERN, D. L. 2012. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc Natl Acad Sci U S A*, 109, 20975-9.
- GILBERT, L., FORREST, H., SCHULTZ, T. & HARVEY, D. J. 1988. Correlations of ultrastructure and pigmentation suggest how genes control development of wing scales of *Heliconius* butterflies. *J. Res. Lepid*, 26, 141-160.
- INDJEIAN, V. B., KINGMAN, G. A., JONES, F. C., GUENTHER, C. A., GRIMWOOD, J., SCHMUTZ, J., MYERS, R. M. & KINGSLEY, D. M. 2016. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell*, 164, 45-56.
- JANSSEN, J. M., MONTEIRO, A. & BRAKEFIELD, P. M. 2001. Correlations between scale structure and pigmentation in butterfly wings. *Evolution & development*, 3, 415-423.
- KOZAK, K. M., WAHLBERG, N., NEILD, A. F., DASMAHAPATRA, K. K., MALLET, J. & JIGGINS, C. D. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *heliconius* butterflies. *Syst Biol*, 64, 505-24.
- LEWINSKY, R. H., JENSEN, T. G., MOLLER, J., STENSALLE, A., OLSEN, J. & TROELSEN, J. T. 2005. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet*, 14, 3945-53.
- LI, Y., LIU, Z., SHI, P. & ZHANG, J. 2010. The hearing gene Prestin unites echolocating bats and whales. *Curr Biol*, 20, R55-6.
- MAROJA, L. S., ALSCHULER, R., MCMILLAN, W. O. & JIGGINS, C. D. 2012. Partial complementarity of the mimetic yellow bar phenotype in *Heliconius* butterflies. *PLoS One*, 7, e48627.
- MARTIN, A., MCCULLOCH, K. J., PATEL, N. H., BRISCOE, A. D., GILBERT, L. E. & REED, R. D. 2014. Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *Evodevo*, 5, 7.
- MARTIN, A. & ORGOGOZO, V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*, 67, 1235-50.
- MAVÁREZ, J., SALAZAR, C. A., BERMINGHAM, E., SALCEDO, C., JIGGINS, C. D. & LINARES, M. 2006. Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441, 868-71.
- MERRILL, R. M., DASMAHAPATRA, K. K., DAVEY, J. W., DELL'AGLIO, D. D., HANLY, J. J., HUBER, B., JIGGINS, C. D., JORON, M., KOZAK, K. M., LLAURENS, V., MARTIN, S. H., MONTGOMERY, S. H., MORRIS, J., NADEAU, N. J., PINHARANDA, A. L., ROSSER, N., THOMPSON, M. J., VANJARI, S., WALLBANK, R. W. & YU, Q. 2015. The diversification of *Heliconius* butterflies: what have we learned in 150 years? *J Evol Biol*, 28, 1417-38.
- PARDO-DIAZ, C., SALAZAR, C., BAXTER, S. W., MEROT, C., FIGUEIREDO-READY, W., JORON, M., MCMILLAN, W. O. & JIGGINS, C. D. 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet*, 8, e1002752.

- PROTAS, M. E., HERSEY, C., KOCHANNEK, D., ZHOU, Y., WILKENS, H., JEFFERY, W. R., ZON, L. I., BOROWSKY, R. & TABIN, C. J. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet*, 38, 107-11.
- PRUD'HOMME, B., GOMPEL, N., ROKAS, A., KASSNER, V. A., WILLIAMS, T. M., YEH, S. D., TRUE, J. R. & CARROLL, S. B. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, 440, 1050-3.
- ROGERS, W. A., SALOMONE, J. R., TACY, D. J., CAMINO, E. M., DAVIS, K. A., REBEIZ, M. & WILLIAMS, T. M. 2013. Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genet*, 9, e1003740.
- SALAZAR, C., BAXTER, S. W., PARDO-DIAZ, C., WU, G., SURRIDGE, A., LINARES, M., BERMINGHAM, E. & JIGGINS, C. D. 2010. Genetic evidence for hybrid trait speciation in heliconius butterflies. *PLoS Genet*, 6, e1000930.
- SALAZAR, C. A., JIGGINS, C. D., ARIAS, C. F., TOBLER, A., BERMINGHAM, E. & LINARES, M. 2005. Hybrid incompatibility is consistent with a hybrid origin of *Heliconius heurippa* Hewitson from its close relatives, *Heliconius cydno* Doubleday and *Heliconius melpomene* Linnaeus. *J Evol Biol*, 18, 247-56.
- STEINMETZ, P. R., URBACH, R., POSNIEN, N., ERIKSSON, J., KOSTYUCHENKO, R. P., BRENA, C., GUY, K., AKAM, M., BUCHER, G. & ARENDT, D. 2010. Six3 demarcates the anterior-most developing brain region in bilaterian animals. *Evodevo*, 1, 14.
- STERN, D. L. & FRANKEL, N. 2013. The structure and evolution of cis-regulatory regions: the shavenbaby story. *Philos Trans R Soc Lond B Biol Sci*, 368, 20130028.
- STERN, D. L. & ORGOGOZO, V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution*, 62, 2155-77.
- SUPPLE, M. A., HINES, H. M., DASMAHAPATRA, K. K., LEWIS, J. J., NIELSEN, D. M., LAVOIE, C., RAY, D. A., SALAZAR, C., MCMILLAN, W. O. & COUNTERMAN, B. A. 2013. Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome research*.
- SWOFFORD, D. L. 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.
- TISHKOFF, S. A., REED, F. A., RANCIARO, A., VOIGHT, B. F., BABBITT, C. C., SILVERMAN, J. S., POWELL, K., MORTENSEN, H. M., HIRBO, J. B., OSMAN, M., IBRAHIM, M., OMAR, S. A., LEMA, G., NYAMBO, T. B., GHORI, J., BUMPSTEAD, S., PRITCHARD, J. K., WRAY, G. A. & DELOUKAS, P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet*, 39, 31-40.
- VAN BELLEGHEM, S. M., RASTAS, P., PAPANICOLAOU, A., MARTIN, S. H., ARIAS, C. F., SUPPLE, M. A., HANLY, J. J., MALLET, J., LEWIS, J. J., HINES, H. M., RUIZ, M., SALAZAR, C., LINARES, M., MOREIRA, G. R. P., JIGGINS, C. D., COUNTERMAN, B. A., MCMILLAN, W. O. & PAPA, R. 2017. Complex modular architecture around a simple toolkit of wing pattern genes. *Nat Ecol Evol*, 1.
- WAGNER, G. P., PAVLICEV, M. & CHEVERUD, J. M. 2007. The road to modularity. *Nat Rev Genet*, 8, 921-31.
- ZHANG, W., DASMAHAPATRA, K. K., MALLET, J., MOREIRA, G. R. & KRONFORST, M. R. 2016a. Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol*, 17, 25.
- ZHANG, W., DASMAHAPATRA, K. K., MALLET, J., MOREIRA, G. R. & KRONFORST, M. R. 2016b. Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol*, 17, 25.

ZHEN, Y., AARDEMA, M. L., MEDINA, E. M., SCHUMER, M. & ANDOLFATTO, P. 2012. Parallel molecular evolution in an herbivore community. *Science*, 337, 1634-7.

Supplementary table 1 Samples of *H. erato*-clade individuals used in this project

Genus	Species	Race	location	country	SRA accession
Heliconius	erato	amalfreda	Brokopondo	Suriname	SAMN05224103
Heliconius	erato	amalfreda	Brokopondo	Suriname	SAMN05224104
Heliconius	erato	amalfreda	Brokopondo	Suriname	SAMN05224208
Heliconius	erato	amalfreda	Brokopondo	Suriname	SAMN05224209
Heliconius	erato	amalfreda	Brokopondo	Suriname	SAMN05224210
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224096
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224097
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224098
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224099
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224192
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224193
Heliconius	erato	chestertonii	Rio Calima	Colombia	SAMN05224194
Heliconius	erato	cyrbia	Balsas	Ecuador	SAMN05224122
Heliconius	erato	cyrbia	Balsas	Ecuador	SAMN05224123
Heliconius	erato	cyrbia	Balsas	Ecuador	SAMN05224124
Heliconius	erato	cyrbia	Balsas	Ecuador	SAMN05224125
Heliconius	erato	demophoon	Gamboa	Panama	SAMN05224182
Heliconius	erato	demophoon	Gamboa	Panama	SAMN05224183
Heliconius	erato	demophoon	Gamboa	Panama	SAMN05224184
Heliconius	erato	demophoon	Gamboa	Panama	SAMN05224185
Heliconius	erato	demophoon	Lago Bayano	Panama	SAMN05224188
Heliconius	erato	demophoon	Lago Bayano	Panama	SAMN05224195
Heliconius	erato	demophoon	Lago Bayano	Panama	SAMN05224196
Heliconius	erato	demophoon	Lago Bayano	Panama	SAMN05224198
Heliconius	erato	demophoon	Lago Bayano	Panama	SAMN05224202
Heliconius	erato	demophoon	Lago Bayano	Panama	SAMN05224203
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224127
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224128
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224154
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224155
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224156
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224157
Heliconius	erato	emma	San Cristobal	Peru	SAMN05224158
Heliconius	erato	erato	Kaw	French Guiana	SAMN05224160
Heliconius	erato	erato	Kaw	French Guiana	SAMN05224161
Heliconius	erato	erato	Kaw	French Guiana	SAMN05224162
Heliconius	erato	erato	Kaw	French Guiana	SAMN05224163
Heliconius	erato	erato	Kaw	French Guiana	SAMN05224164
Heliconius	erato	erato	Kaw	French Guiana	SAMN05224174
Heliconius	erato	etylus	Huamboya	Ecuador	SAMN05224110
Heliconius	erato	etylus	Huamboya	Ecuador	SAMN05224111
Heliconius	erato	etylus	Huamboya	Ecuador	SAMN05224112
Heliconius	erato	etylus	Huamboya	Ecuador	SAMN05224113
Heliconius	erato	etylus	Huamboya	Ecuador	SAMN05224114
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224126
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224148
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224149
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224150
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224151
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224152
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224172
Heliconius	erato	favorinus	Urahuasha	Peru	SAMN05224173
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224153
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224159
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224165
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224166
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224167
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224175

Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224176
Heliconius	erato	hydara	Kaw	French Guiana	SAMN05224177
Heliconius	erato	hydara	Lago Bayano	Panama	SAMN05224189
Heliconius	erato	hydara	Lago Bayano	Panama	SAMN05224190
Heliconius	erato	hydara	Lago Bayano	Panama	SAMN05224197
Heliconius	erato	hydara	Lago Bayano	Panama	SAMN05224200
Heliconius	erato	hydara	Lago Bayano	Panama	SAMN05224201
Heliconius	erato	hydara	Lago Bayano	Panama	SAMN05224191
Heliconius	erato	lativitta	Ahuano	Ecuador	SAMN05224101
Heliconius	erato	lativitta	Ahuano	Ecuador	SAMN05224137
Heliconius	erato	lativitta	Ahuano	Ecuador	SAMN05224138
Heliconius	erato	lativitta	Ahuano	Ecuador	SAMN05224139
Heliconius	erato	lativitta	Jondachi	Ecuador	SAMN05224140
Heliconius	erato	notabilis	Huamboya	Ecuador	SAMN05224105
Heliconius	erato	notabilis	Huamboya	Ecuador	SAMN05224106
Heliconius	erato	notabilis	Huamboya	Ecuador	SAMN05224107
Heliconius	erato	notabilis	Huamboya	Ecuador	SAMN05224108
Heliconius	erato	notabilis	Huamboya	Ecuador	SAMN05224109
Heliconius	erato	notabilis	Mera	Ecuador	SAMN05224100
Heliconius	erato	notabilis	Mera	Ecuador	SAMN05224178
Heliconius	erato	notabilis	Mera	Ecuador	SAMN05224179
Heliconius	erato	notabilis	Mera	Ecuador	SAMN05224180
Heliconius	erato	notabilis	Mera	Ecuador	SAMN05224181
Heliconius	erato	petiverana	Campeche	Mexico	SAMN05224115
Heliconius	erato	petiverana	Campeche	Mexico	SAMN05224116
Heliconius	erato	petiverana	Campeche	Mexico	SAMN05224117
Heliconius	erato	petiverana	Campeche	Mexico	SAMN05224118
Heliconius	erato	petiverana	Campeche	Mexico	SAMN05224119
Heliconius	erato	cruentus	Puerto Vallarta	Mexico	SAMN05224147
Heliconius	erato	phyllis	Samaipata	Bolivia	SAMN05224204
Heliconius	erato	phyllis	Samaipata	Bolivia	SAMN05224205
Heliconius	erato	phyllis	Samaipata	Bolivia	SAMN05224206
Heliconius	erato	phyllis	Samaipata	Bolivia	SAMN05224207
Heliconius	erato	venus	Queremal	Colombia	SAMN05224141
Heliconius	erato	venus	Queremal	Colombia	SAMN05224142
Heliconius	erato	venus	Queremal	Colombia	SAMN05224143
Heliconius	erato	venus	Queremal	Colombia	SAMN05224144
Heliconius	erato	venus	Queremal	Colombia	SAMN05224145
Heliconius	himera	-	Vilcabamba	Ecuador	SAMN05224132
Heliconius	himera	-	Vilcabamba	Ecuador	SAMN05224133
Heliconius	himera	-	Vilcabamba	Ecuador	SAMN05224134
Heliconius	himera	-	Vilcabamba	Ecuador	SAMN05224135
Heliconius	himera	-	Vilcabamba	Ecuador	SAMN05224136
Genus	Species	Race	location	country	SRA accession

Supplementary table 2: Samples of *H. melpomene*-clade individuals used in this project

Genus	Species	Race	rays	dennis	band	Sample Code	Location
<i>Heliconius</i>	<i>aoede</i>		0	0	0	JM-09-347	Michaela Bastida, Peru
<i>Heliconius</i>	<i>cydno</i>	<i>alitha</i>	0	0	0	CAM009999	Insectary reared
<i>Heliconius</i>	<i>cydno</i>	<i>alitha</i>	0	0	0	CAM008509	Ecuador, Pichincha
<i>Heliconius</i>	<i>cydno</i>	<i>alitha</i>	0	0	0	CAM008517	Ecuador, Pichincha
<i>Heliconius</i>	<i>cydno</i>	<i>chioneus</i>	0	0	0	CAM000553	Panama
<i>Heliconius</i>	<i>cydno</i>	<i>chioneus</i>	0	0	0	CAM000560	Panama
<i>Heliconius</i>	<i>cydno</i>	<i>chioneus</i>	0	0	0	CAM000564	Panama
<i>Heliconius</i>	<i>cydno</i>	<i>chioneus</i>	0	0	0	CAM000565	Panama
<i>Heliconius</i>	<i>cydno</i>	<i>cordula</i>	0	0	0	STRI_007	San Cristobal, Venezuela
<i>Heliconius</i>	<i>cydno</i>	<i>cordula</i>	0	0	0	M2258	San Cristobal, Venezuela
<i>Heliconius</i>	<i>cydno</i>	<i>cordula</i>	0	0	0	M2255	San Cristobal, Venezuela
<i>Heliconius</i>	<i>cydno</i>	<i>cordula</i>	0	0	0	M2253	San Cristobal, Venezuela
<i>Heliconius</i>	<i>cydno</i>	<i>cordula</i>	0	0	0	M2157	San Cristobal, Venezuela
<i>Heliconius</i>	<i>cydno</i>	<i>cydnides</i>	0	0	0	CS002017	Marsella-Risaralda, Colombia
<i>Heliconius</i>	<i>cydno</i>	<i>cydnides</i>	0	0	0	CS002018	Marsella-Risaralda, Colombia
<i>Heliconius</i>	<i>cydno</i>	<i>hermogenes</i>	0	0	0	CS003278	Santafé de los guaduales,
<i>Heliconius</i>	<i>cydno</i>	<i>weymeri f.</i>	0	0	0	CS002529	Helechaux, Colombia
<i>Heliconius</i>	<i>cydno</i>	<i>weymeri f.</i>	0	0	0	CS001690	Helechaux, Colombia
<i>Heliconius</i>	<i>cydno</i>	<i>zelinde</i>	0	0	0	CS002262	Ladrilleros, Colombia
<i>Heliconius</i>	<i>doris</i>		0	0	0	JM-02-1939	Km-5 Shapaja-Chazuta, Peru
<i>Heliconius</i>	<i>doris</i>	<i>doris (blue)</i>	0	0	0	CAM008684	San Martin, Peru
<i>Heliconius</i>	<i>doris</i>	<i>delila (red)</i>	0	0	0	CAM008697	San Martin, Peru
<i>Heliconius</i>	<i>elevatus</i>	<i>bari</i>	1	1	0	MJ09-4037	Patawa Laie, French Guiana
<i>Heliconius</i>	<i>elevatus</i>	<i>bari</i>	1	1	0	MJ09-4056	Patawa Laie, French Guiana
<i>Heliconius</i>	<i>elevatus</i>	<i>bari</i>	1	1	0	MJ09-4094	Patawa Laie, French Guiana
<i>Heliconius</i>	<i>elevatus</i>		1	1	0	JM-09-343	Km-103.1 Tarapoto-
<i>Heliconius</i>	<i>elevatus</i>		1	1	0	JM-09-118	Munichis, Peru
<i>Heliconius</i>	<i>elevatus</i>		1	1	0	JM-09-163	Km-75 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>elevatus</i>		1	1	0	JM-09-270	Michaela Bastida, Peru
<i>Heliconius</i>	<i>elevatus</i>		1	1	0	JM-09-302	Km-17.2 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>elevatus</i>		1	1	0	BC_0408	Pimpiala, Ecuador
<i>Heliconius</i>	<i>ethilla</i>	<i>aerotome</i>	0	0	0	JM-09-63	Km-8 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>ethilla</i>		0	0	0	JM-09-49	Km-17.2 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>ethilla</i>	<i>aerotome</i>	0	0	0	JM-09-67	Urahuasha, Peru
<i>Heliconius</i>	<i>ethilla</i>		0	0	0	JM-09-62	Urahuasha, Peru
<i>Heliconius</i>	<i>ethilla</i>		0	0	0	JM-09-66	Urahuasha, Peru
<i>Heliconius</i>	<i>hecale</i>		0	0	0	JM-02-1326	Km-7.2 Pongo-Barranquita,
<i>Heliconius</i>	<i>hecale</i>		0	0	0	JM-09-272	Michaela Bastida, Peru
<i>Heliconius</i>	<i>hecale</i>		0	0	0	JM-09-164	Km-75 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>hecale</i>		0	0	0	STRI_001	Gamboa, Panama
<i>Heliconius</i>	<i>hecale</i>	<i>felix</i>	0	0	0	JM-09-345	Km-103.1 Tarapoto-
<i>Heliconius</i>	<i>hecale</i>	<i>felix</i>	0	0	0	JM-09-273	Peru
<i>Heliconius</i>	<i>hecuba</i>	<i>flava</i>	0	0	0	CAM008550	Sucumbios, near La Bonita,
<i>Heliconius</i>	<i>heurippa</i>		0	0	1	CH9-	Buenavista, Colombia
<i>Heliconius</i>	<i>heurippa</i>		0	0	1	STRI_002	Buenavista, Colombia
<i>Heliconius</i>	<i>hierax</i>		0	0	0	CAM008149	Napo, Ecuador
<i>Heliconius</i>	<i>ismenius</i>		0	0	0	STRI_003	Gamboa, Panama
<i>Heliconius</i>	<i>ismenius</i>	<i>metaphorus</i>	0	0	0	CAM009995	Western Ecuador, insectary
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	CAM009998	Insectary reared
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-11-572	Michaela Bastida, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-11-569	Michaela Bastida, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-09-246	Km-103.1 Tarapoto-
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-09-267	Km-103.1 Tarapoto-
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-09-268	Idea Religiosa, Munichis, Peru

<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-09-357	Km-103.1 Tarapoto-Yurimaguas, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	JM-09-112	Munichis, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	1048-143N20	Insectary reared, Fosmid
<i>Heliconius</i>	<i>melpomene</i>	<i>aglaope</i>	1	1	0	1048-3N15	Insectary reared, Fosmid
<i>Heliconius</i>	<i>melpomene</i>	<i>amandus</i>	0	0	1	CS002221	Angostura, Bolivia
<i>Heliconius</i>	<i>melpomene</i>	<i>amandus</i>	0	0	1	CS002228	Angostura, Bolivia
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	CAM009997	Insectary reared
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-09-216	Puente Serranoyacu, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-11-160	Rio Shilcayo, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-11-293	Urahuasha, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-09-332	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-09-333	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-09-75	Km-8 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>melpomene</i>	<i>amaryllis</i>	0	0	1	JM-09-79	Km-8 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>melpomene</i>	<i>cythera</i>	0	0	1	CAM002856	Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>cythera</i>	0	0	1	CAM002857	Pichincha Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>bellula</i>	0	0	1	CS000228	Putumayo, Colombia
<i>Heliconius</i>	<i>melpomene</i>	<i>bellula</i>	0	0	1	CS000231	Putumayo, Colombia
<i>Heliconius</i>	<i>melpomene</i>	<i>ecuadorensis</i>	1	1	0	CAM009117	Zamora, Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>ecuadorensis</i>	1	1	0	CAM009121	Zamora, Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>malleti</i> x	0	1	0	CAM016042	Ecuador (Col Mariscal I)
<i>Heliconius</i>	<i>melpomene</i>	<i>malleti</i>	1	1	0	CAM016550	Y de Misahuallí, Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>malleti</i>	1	1	0	CAM017162	Y de Misahuallí, Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	STRI_WOM_0	Darién, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	STRI_WOM_0	Darién, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	CAM018097	Darién, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	CAM018038	Darién, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	STRI_006	Colombia
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	CAM013435	French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	CAM009315	French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	CAM009316	French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	CAM009317	French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>melpomene</i>	0	0	1	gen_ref	Darién, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>mixed</i>	0	0	1	AEHM-19L14	Insectary reared, BAC library
<i>Heliconius</i>	<i>melpomene</i>	<i>meriana</i>	0	1	0	CAM013819	Neuve Wakapo, French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>meriana</i>	0	1	0	CAM013715	Neuve Wakapo, French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>plesseni</i>	0	0	1	CAM009156	Tungurahua, Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>plesseni</i>	0	0	1	CAM016293	Pindo-Mirador, Ecuador
<i>Heliconius</i>	<i>melpomene</i>	<i>rosina</i>	0	0	1	CAM009996	Insectary reared
<i>Heliconius</i>	<i>melpomene</i>	<i>rosina</i>	0	0	1	CAM002071	Gamboa, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>rosina</i>	0	0	1	CAM000531	Gamboa, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>rosina</i>	0	0	1	CAM000533	Gamboa, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>rosina</i>	0	0	1	CAM000546	Gamboa, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>thelxiopeia</i>	1	1	0	CAM013566	Neuve Wakapo, French Guiana
<i>Heliconius</i>	<i>melpomene</i>	<i>vulcanus</i>	0	0	1	CAM014632	Darién, Panama
<i>Heliconius</i>	<i>melpomene</i>	<i>vulcanus</i>	0	0	1	CS000519	Rio Bravo-Calima, Colombia
<i>Heliconius</i>	<i>numata</i>	<i>bicoloratus</i>	0	0	0	MJ05_123	Fundo Biodiversidad km19
<i>Heliconius</i>	<i>numata</i>	<i>bicoloratus</i>	0	0	0	JM-05-1116	Puente Rio Serranoyacu Rioja
<i>Heliconius</i>	<i>numata</i>	<i>arcuella</i>	0	0	0	JM-05-1277	Km-26 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>numata</i>	<i>silvana</i>	0	0	0	JM-09-364	Km-18 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>numata</i>	<i>tarapotensis</i>	0	0	0	JM-05-1358	Km-8 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>pachinus</i>		0	0	0	CAM008020	Chiriqui Panama
<i>Heliconius</i>	<i>pachinus</i>		0	0	0	CAM008035	Chiriqui Panama
<i>Heliconius</i>	<i>pardalinus</i>	<i>sergestus</i>	0	0	0	JM-09-326	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>sergestus</i>	0	0	0	JM-09-202	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>sergestus</i>	0	0	0	JM-09-201	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>sergestus</i>	0	0	0	JM-09-209	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>sergestus</i>	0	0	0	JM-09-210	Tarapoto - Urahuasha trail, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>ssp. nov.</i>	0	0	0	JM-09-372	Caño Tushmo, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>ssp. nov.</i>	0	0	0	JM-09-373	Caño Tushmo, Peru

<i>Heliconius</i>	<i>pardalinus</i>	<i>ssp. nov.</i>	0	0	0	JM-09-371	Caño Tushmo, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>ssp. nov.</i>	0	0	0	JM-09-374	Caño Tushmo, Peru
<i>Heliconius</i>	<i>pardalinus</i>	<i>ssp. nov.</i>	0	0	0	JM-09-387	Caño Tushmo, Peru
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	1	0	0	CAM009223	Tungurahua, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	1	0	0	BC_0406	Tungurahua, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	1	0	0	CAM008520	Tungurahua, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	1	0	0	CAM008523	Tungurahua, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>florencia</i>	1	1	0	CS002403	Quebrada Doraditas, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>florencia</i>	1	1	0	CS002406	Quebrada Doraditas, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>florencia</i>	1	1	0	CS002407	Quebrada Doraditas, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>florencia</i>	1	1	0	CS002410	Quebrada Doraditas, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>linaresi</i>	0	0	0	CS002234	Puerto Amor, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>linaresi</i>	0	0	0	CS002409	Puerto Amor, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>linaresi</i>	0	0	0	CS002434	Puerto Amor, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>linaresi</i>	0	0	0	CS002435	Puerto Amor, Colombia
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	JM-09-312	Km-18 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	JM-09-57	Tarapoto-Yurimaguas, Peru
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	JM-09-86	Tarapoto-Yurimaguas, Peru
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	JM-09-313	Km-18 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	CAM008624	Km-15 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	CAM008628	Km-15 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>timareta</i>	<i>thelxinoe</i>	0	0	1	CAM008631	Km-15 Tarapoto-Yurimaguas,
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	0	0	0	CAM009178	El Topo, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	0	0	0	BC_0407	Tungurahua, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	0	0	0	CAM008533	Tungurahua, Ecuador
<i>Heliconius</i>	<i>timareta</i>	<i>timareta</i>	0	0	0	CAM009184	Tungurahua, Ecuador
<i>Heliconius</i>	<i>wallacei</i>		0	0	0	JM-04-200	Pucallpa, Peru
<i>Heliconius</i>	<i>xanthocles</i>		0	0	0	CAM009106	Zamora-Chinchi, Ecuador

Supplementary figure S2.1: topologies for *twisst* analysis of relationships between *H. melpomene* and *H. besckei*.

topo0

```

      /-H_melpomene_band
      /-|
      /-| \-H_melpomene_DennisRay
      ||
      || \-H_besckei
      -|
      -| \-Silvaniforms
      |
      | \-H_cydno

```

topo1

```

      /-H_melpomene_band
      /-|
      /-| \-H_melpomene_DennisRay
      ||
      || \-Silvaniforms
      -|
      -| \-H_besckei
      |
      | \-H_cydno

```

topo2

```

      /-H_melpomene_band
      /-|
      /-| \-H_melpomene_DennisRay
      ||
      || \-H_cydno
      -|
      -| \-H_besckei
      |
      | \-Silvaniforms

```

topo3

```

      /-H_melpomene_band
      /-|
      /-| \-H_besckei
      ||
      || \-H_melpomene_DennisRay
      -|
      -| \-Silvaniforms
      |
      | \-H_cydno

```

topo4

```

      /-H_melpomene_band
      /-|
      /-| \-H_besckei
      ||
      || \-Silvaniforms
      -|
      -| \-H_melpomene_DennisRay
      |
      | \-H_cydno

```

topo5

```

      /-H_melpomene_band
      /-|
      /-| \-H_besckei
      ||
      || \-H_cydno
      -|
      -| \-H_melpomene_DennisRay
      |
      | \-Silvaniforms

```

topo6

```

      /-H_melpomene_band
      /-|
      /-| \-Silvaniforms
      ||
      || \-H_melpomene_DennisRay
      -|
      -| \-H_besckei
      |
      | \-H_cydno

```

topo7

```

      /-H_melpomene_band
      /-|
      /-| \-Silvaniforms
      ||
      || \-H_besckei
      -|
      -| \-H_melpomene_DennisRay
      |
      | \-H_cydno

```

topo8

```

      /-H_melpomene_band
      /-|
      /-| \-Silvaniforms
      ||
      || \-H_cydno
      -|
      -| \-H_melpomene_DennisRay
      |
      | \-H_besckei

```

topo9

```

      /-H_melpomene_band
      /-|
      /-| \-H_cydno
      ||
      || \-H_melpomene_DennisRay
      -|
      -| \-H_besckei
      |
      | \-Silvaniforms

```

topo10

```

      /-H_melpomene_band
      /-|
      /-| \-H_cydno
      ||
      || \-H_besckei
      -|
      -| \-H_melpomene_DennisRay
      |
      | \-Silvaniforms

```

topo11

```

      /-H_melpomene_band
      /-|
      /-| \-H_cydno
      ||
      || \-Silvaniforms
      -|
      -| \-H_melpomene_DennisRay
      |
      | \-H_besckei

```

topo12

```

      /-H_melpomene_band
      /-|
      || /-H_melpomene_DennisRay
      || \-|
      -| \-H_besckei
      |
      | \-Silvaniforms
      |
      | \-H_cydno

```

topo13

```

      /-H_melpomene_band
      /-|
      || /-H_melpomene_DennisRay
      || \-|
      -| \-Silvaniforms
      |
      | \-H_besckei
      |
      | \-H_cydno

```

topo14

```

      /-H_melpomene_band
      /-|
      || /-H_melpomene_DennisRay
      || \-|
      -| \-H_cydno
      |
      | \-H_besckei
      |
      | \-Silvaniforms

```


Supplementary figure S2.2: topologies for *twisst* analysis of relationships between *H. melpomene*, *H. timareta linaresi* and *H. heurippa*.

topo0

```

      /-H_melpomene_band
      /-|
      /-| \-H_cydno
      ||
      | \-H_heurippa
      -|
      |--H_timareta_linaresi
      |
      \-H_melpomene_DennisRay

```

topo1

```

      /-H_melpomene_band
      /-|
      /-| \-H_cydno
      ||
      | \-H_timareta_linaresi
      -|
      |--H_heurippa
      |
      \-H_melpomene_DennisRay

```

topo2

```

      /-H_melpomene_band
      /-|
      /-| \-H_cydno
      ||
      | \-H_melpomene_DennisRay
      -|
      |--H_heurippa
      |
      \-H_timareta_linaresi

```

topo3

```

      /-H_melpomene_band
      /-|
      /-| \-H_heurippa
      ||
      | \-H_cydno
      -|
      |--H_timareta_linaresi
      |
      \-H_melpomene_DennisRay

```

topo4

```

      /-H_melpomene_band
      /-|
      /-| \-H_heurippa
      ||
      | \-H_timareta_linaresi
      -|
      |--H_cydno
      |
      \-H_melpomene_DennisRay

```

topo5

```

      /-H_melpomene_band
      /-|
      /-| \-H_heurippa
      ||
      | \-H_melpomene_DennisRay
      -|
      |--H_cydno
      |
      \-H_timareta_linaresi

```

topo6

```

      /-H_melpomene_band
      /-|
      /-| \-H_timareta_linaresi
      ||
      | \-H_cydno
      -|
      |--H_heurippa
      |
      \-H_melpomene_DennisRay

```

topo7

```

      /-H_melpomene_band
      /-|
      /-| \-H_timareta_linaresi
      ||
      | \-H_heurippa
      -|
      |--H_cydno
      |
      \-H_melpomene_DennisRay

```

topo8

```

      /-H_melpomene_band
      /-|
      /-| \-H_timareta_linaresi
      ||
      | \-H_melpomene_DennisRay
      -|
      |--H_cydno
      |
      \-H_heurippa

```

topo9

```

      /-H_melpomene_band
      /-|
      /-| \-H_melpomene_DennisRay
      ||
      | \-H_cydno
      -|
      |--H_heurippa
      |
      \-H_timareta_linaresi

```

topo10

```

      /-H_melpomene_band
      /-|
      /-| \-H_melpomene_DennisRay
      ||
      | \-H_heurippa
      -|
      |--H_cydno
      |
      \-H_timareta_linaresi

```

topo11

```

      /-H_melpomene_band
      /-|
      /-| \-H_melpomene_DennisRay
      ||
      | \-H_timareta_linaresi
      -|
      |--H_cydno
      |
      \-H_heurippa

```

topo12

```

      /-H_melpomene_band
      /-|
      || /-H_cydno
      | \-|
      -| \-H_heurippa
      |
      |--H_timareta_linaresi
      |
      \-H_melpomene_DennisRay

```

topo13

```

      /-H_melpomene_band
      /-|
      || /-H_cydno
      | \-|
      -| \-H_timareta_linaresi
      |
      |--H_heurippa
      |
      \-H_melpomene_DennisRay

```

topo14

```

      /-H_melpomene_band
      /-|
      || /-H_cydno
      | \-|
      -| \-H_melpomene_DennisRay
      |
      |--H_heurippa
      |
      \-H_timareta_linaresi

```

Chapter 3

Pattern differences across a hybrid zone; using RNA-seq to characterise molecular differences between mimetic pattern morphs

ABSTRACT

The wing patterns of butterflies and moths are diverse, and provide striking examples of evolution by natural selection. The gene *cortex* has been linked with wing pattern development in several lepidopteran species including *Heliconius*, where candidate regulatory regions for yellow pattern elements have been annotated in multiple species. I investigated gene expression differences in two morphs of *H. melpomene* and *H. erato* from either side of a hybrid zone that vary only in the presence or absence of a yellow pattern element, in order to determine a role for candidate genes at the yellow pattern locus. In *H. melpomene* the gene *cortex* was upregulated in the larval wing discs of the black morph, whereas in *H. erato* it was upregulated in the larval wing discs of the yellow morph. In pupal wings, *washout* was differentially expressed, again in the opposite pattern in the two species, suggesting the same locus is responsible for convergent pattern modification, but by a different mechanism. This has offered insight into the processes by which the causative agents at this locus are differentially regulated to steer pattern development.

I: INTRODUCTION

The utilization of ‘hotspots’ of evolution has recently become a recurring theme of evolutionary biology (Martin and Orgogozo, 2013). *Heliconius* is no exception; genomic analysis has revealed that three loci of major effect account for most of the genetic diversity within species (Nadeau et al., 2013, Van Belleghem et al., 2017), and that these

same loci are hotspots for pattern evolution in several species. For example, the yellow pattern locus has been linked to wing pattern in *H. melpomene* and *H. erato*, but also variation in all pattern elements in *H. numata* (Joron et al., 2006, Nadeau et al., 2014). The homologous locus has also been linked to the industrial melanism phenotype in the peppered moth *Biston betularia*, as well as to the Bigeye family of mutants in *Bicyclus anynana* and melanisation patterns in the silk moth *Bombyx mori* (Van't Hof et al., 2016, Beldade et al., 2009, Ito et al., 2016). Repeated evidence for functional evolution at this locus implies that it may contain one or more genes that play a critical role in the gene regulatory network that leads to the development of wing pattern in Lepidoptera, but unlike other such loci, we do not have extensive correlative expression data or functional validation of this role.

Molecular mechanisms of yellow pattern elements

Different colour patches on *Heliconius* wings differ in both scale colour and structure (Gilbert et al (1988); Type I scales either contain the yellow pigment 3-hydroxykynurenine 3'OHK or no pigment, in which case they are white due to structural diffraction. Yellow fated regions can be distinguished with low-magnification stereomicroscopy as early as 72 h of pupal development because Type I scales undergo laminar extension before other scale types, causing a visible difference in density. This indicates heterochrony in the timing of differentiation between different scale cell types.

H. melpomene

In both mimetic species, yellow pattern elements are controlled by major effect loci. Linkage and association analyses in *H. melpomene* have found three pattern elements controlled by a locus on chromosome 15: the hindwing Yellow bar (*Yb*), the white hindwing margin patterns (*Sb*), and presence of yellow patterns on the forewing (*N*). Other aspects of these yellow patterns are controlled by unlinked loci: Yellow vs white colour (*K*), is linked to chromosome 1, and variation in the shape of the forewing band element is controlled by *WntA* (Figure 3.1) (Martin and Reed, 2014, Martin et al., 2012, Joron et al., 2006, Sheppard et al., 1985).

Nadeau et al showed that a gene named *cortex* is implicated at this locus. Fine mapping and association studies using populations of resequenced individuals showed peaks of association around the gene *cortex* (Ferguson et al., 2010, Nadeau et al., 2012).

Microarray expression analyses on different pattern races showed differential expression of one gene, *cortex*, in association with wing pattern. *in situ* hybridization of *cortex* in late larval wings of *H. m. rosina* showed expression in negative association with the yellow bar, implying an inhibitory effect of *cortex* on differentiation into the yellow Type I scale cell. Some exons of *cortex* were absent in some pattern forms, and that splice variation of *cortex* correlates with pattern (Nadeau et al., 2016).

I will refer to this whole chromosome 15 yellow colour pattern locus as “the *cortex* locus” here, albeit with the acknowledgement that other genes in this region may also be implicated in functional control of colour pattern.

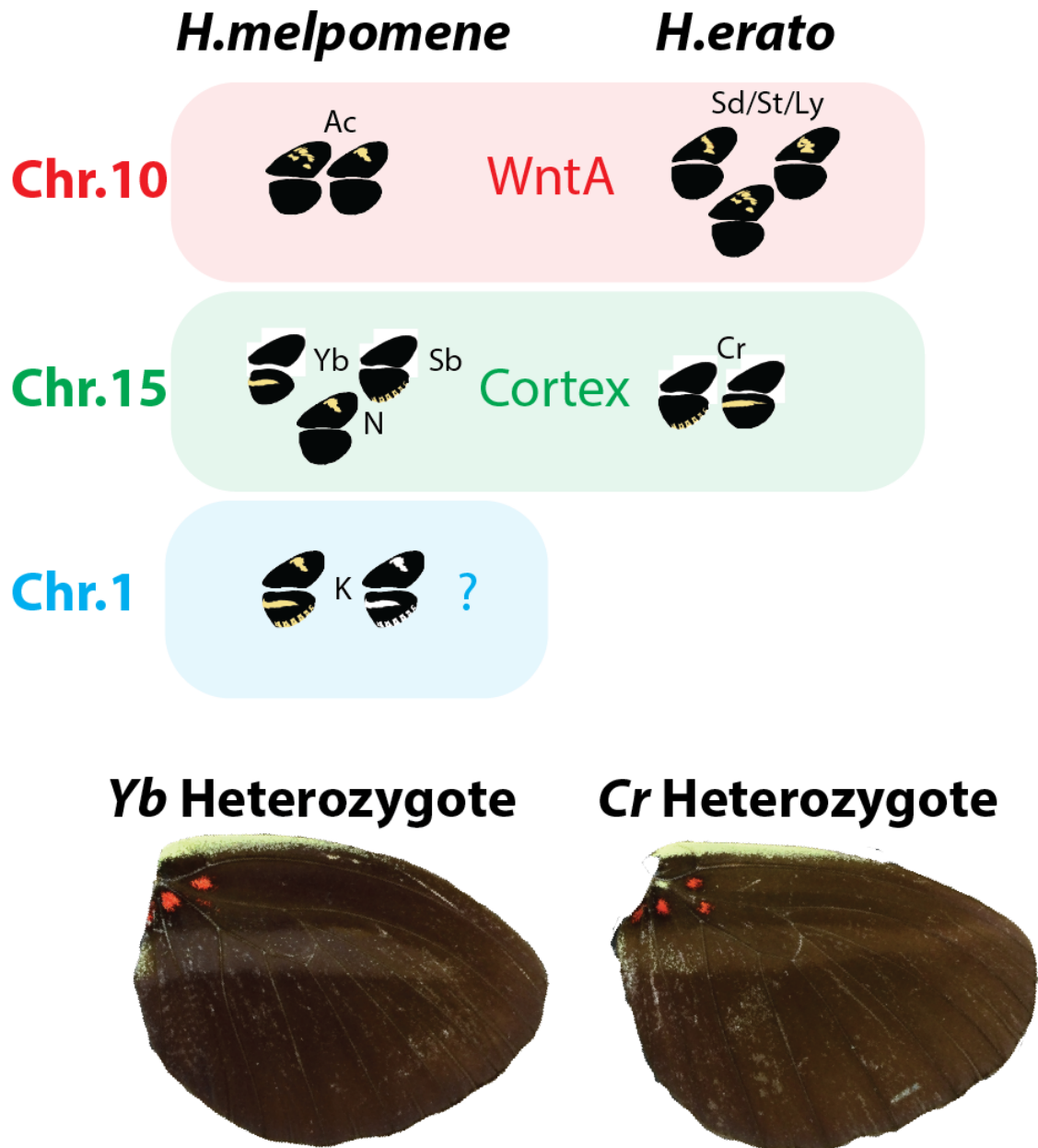


Figure 3.1: Graphical summary of yellow and white pattern loci described by Sheppard et al. On the left are the loci described in *H. melpomene*, and on the right the loci described in *H. erato*. Loci were given different names in each species, indicated in black. These loci were later mapped to the same chromosomal regions, and in the case of the chromosome 10- and 15-linked loci, candidate functional genes have been proposed. Below, the heterozygote forms of *Yb* and *Cr*, with “shadow” scales on the ventral surface.

H. erato

The same locus also controls convergent mimetic patterns in *H. erato*. Linkage analysis in *H. erato* has shown that *Cr*, the yellow bar locus in *H. erato*, is also linked to *cortex*. However there are differences between the two species, as this locus is not associated with the forewing band colour in *H. erato* (Figure 3.1) (Joron et al., 2006, Van Belleghem et al., 2017). In addition, there is evidence for two independent origins of the yellow bar within the *H. erato* lineage. *Cr* alleles from east and west of the Andes do not complement in hybrids, and association mapping indicates that eastern races have an association centred on *cortex*, whereas western races (*H. e. hydara* vs *H. e. demophoon*), show association in a second window approx. 100 kb from the first, including coding portions of the gene *parn* (Maroja et al., 2012, Van Belleghem et al., 2017)

Heterozygotes

Unlike red pattern alleles, where the presence of the element is dominant, with yellow pattern alleles at *cortex* the absence of the element is dominant. However, heterozygotes have black scales in the yellow band region that are structurally distinct, and sometimes include a few yellow scales – this is visible on close examination (allowing one to visually genotype heterozygotes), but this is not likely to be identifiable by bird predators so heterozygote butterflies are likely to be functionally identical to homozygote +/- butterflies in terms of selection (Figure 3.1).

In summary, fine-scale association mapping in combination with expression data, has implicated the gene *cortex* as being the causative gene at the yellow pattern locus, but this remains a puzzling candidate. This gene is expressed in the germline of *Drosophila*, and is a member of a family of cell cycle regulators (Nadeau et al.). At other patterning loci, there is now functional evidence from pharmacological perturbation and CRISPR-cas9 mutagenesis, but this is not yet the case for *cortex*. It is unclear how a cell cycle regulator can modulate a gene regulatory network to modify a cell fate decision. Additionally, the nearby genes *domeless* and *washout* have been implicated as the functional genes for the *Bigeye* family of mutants in *B. anynana* (Saenko et al., 2010). Here, I examine transcriptional differences in wings across a hybrid zone in the Darien, Panama, where

both *H. melpomene* and *H. erato* vary in the presence or absence of a yellow bar on the hindwing.

The Panama hybrid zone – origins and evolutionary history

Patterns of selection on these phenotypes are also well understood, particularly for hindwing yellow bar pattern found in Panama and western Colombia. Distinct geographic populations may hybridise freely in areas of geographic contact known as hybrid zones and these regions can provide estimates of the strength of selection (Mallet, 1993). In *Heliconius* there are many stable hybrid zones of varying width and breadth throughout the Americas. A well-described parallel hybrid zone occurs in eastern Panama. A broad, three-way hybrid zone between forms that differ primarily in their hindwing yellow bar stretches across Eastern Panama and the Darien into the Caribbean and Pacific coasts of Colombia (Figure 3.2). The only pattern difference across this hybrid zone, the hindwing yellow band, is therefore linked to variation at the *cortex* locus.

This hybrid zone is very wide, suggesting relatively weak selection (Mallet, 1986). However, if the alleles were not under selection, one would predict a gradual increase in cline width over time, but sampling over an extended time period (1983, 2000, 2015), (Blum, 2002, Mallet, 1986)(Thurman et al., unpublished) shows this has not been the case, indicating persistent selection on the yellow band.

Mallet observed that because of the dominance of the black *-/-* allele, there is no heterozygote disadvantage to heterozygotes in the eastern part of the hybrid zone. This led him to predict that the position of the hybrid zone would be unstable due to dominance drive, and so the centre of the hybrid zone would move westward at a predictable rate – this prediction was later confirmed by Blum et al (2000) and Thurman (unpublished), who have shown that the hybrid zone is moving westward at the rate of about 2.6 km per year (Figure 3.3). In summary, the yellow band is an adaptive phenotype on which there is good evidence for the action of natural selection.

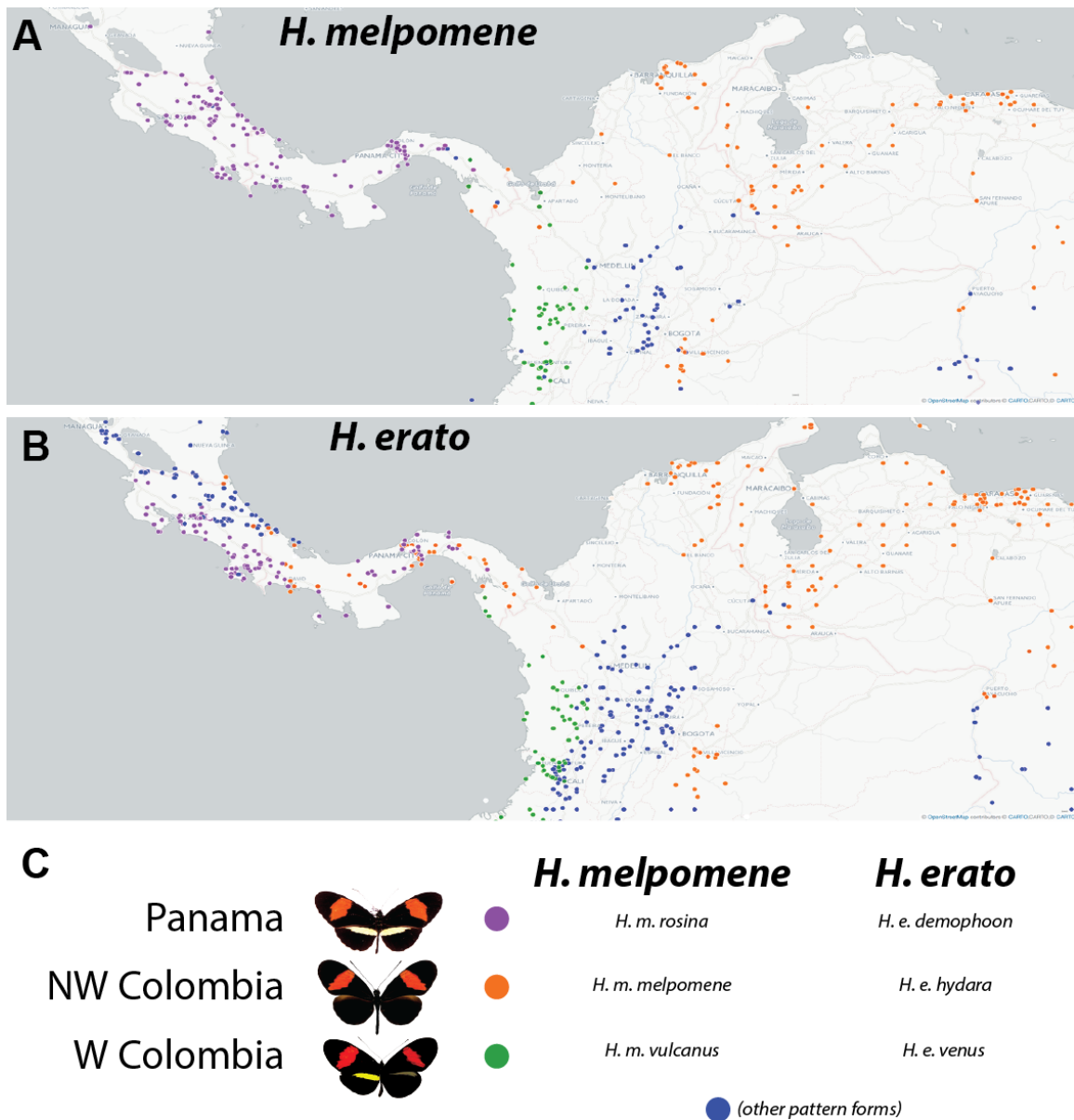


Figure 3.2: Distribution of pattern forms of *H. melpomene* and *H. erato* in Panama and Colombia. Purple dots indicate the yellow pattern form (*H. e. demophoon* and *H. m. rosina*), orange dots indicate the black pattern form (*H. e. hydara* and *H. m. melpomene*), and green dots indicate the dorsal-only yellow pattern form (*H. e. venus* and *H. m. vulcanus*). **A** shows records of *H. melpomene*, **B** shows records of *H. erato*, and **C** shows the pattern forms and race names. Note that the west Colombian forms *H. melpomene vulcanus* and *H. erato venus* have a yellow bar on the ventral surface of their wings but no yellow bar on the dorsal surface. Differences between dorsal and ventral patterns are

uncommon in *Heliconius*. While the distributions are largely similar between the two species, their distributions are not perfectly congruent – for example, *H. erato hydara* extends further west into Panama than *H. melpomene melpomene*. (Note; *H. erato* in Panama was previously described as *H. e. petiverana*, and named as such in many papers through the 1980s-2010s. It has recently been decided that this name was misapplied, and so the Panamanian race is hereafter referred to as *H. e. demophoon*. Here, I will preferentially refer to all races by their wing pattern and geographic origin, and not their binomial designation).

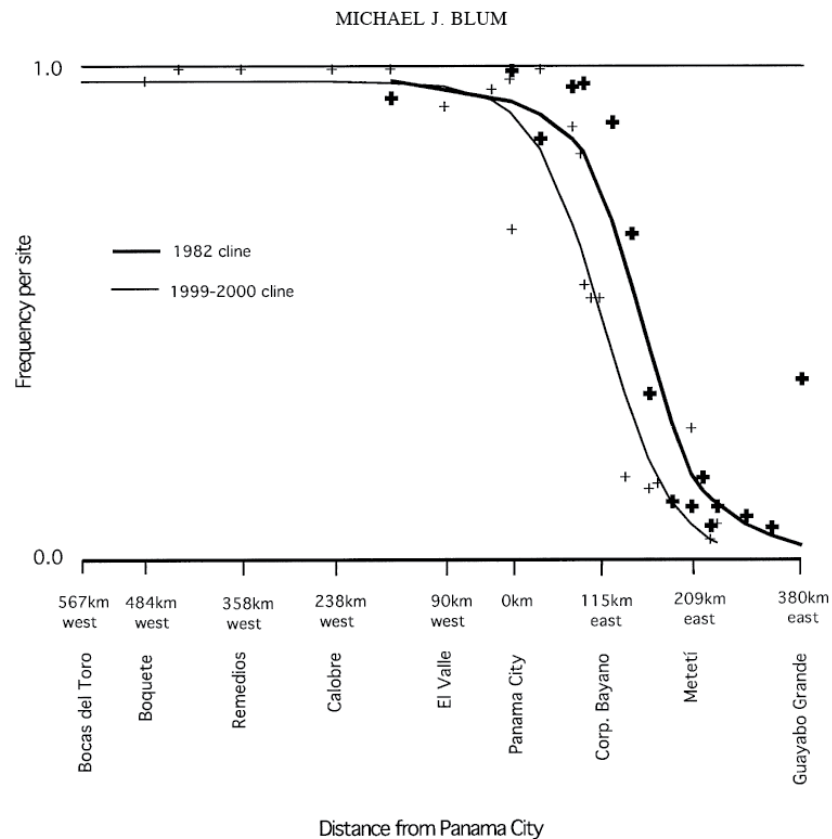


Figure 3.3: Movement of hybrid zone between 1982 and 2000 (Taken from Blum et al 2000). Additional sampling carried out in 2015 has illustrated that the hybrid zone has continued to move west at approximately the same rate (Thurman, pers comm.). Butterflies utilized in this project were collected at Pipeline Road in Soberanía National Park, Gamboa, Panama (around 30 km northwest of Panama City), and from Puerto Lara, near Metetí, (around 209 km east of Panama City).

Project aims

The yellow bar phenotype is therefore an adaptive phenotype for which a locus has been identified, a candidate gene proposed, and candidate regulatory regions annotated. However, there is currently no spatial expression data in either *H. melpomene* or *H. erato* directly comparing races from both sides of the Darien hybrid zone. Aside from *cortex*, there is also evidence from other butterflies that the JAK-STAT pathway ligand *domeless* and the actin-binding protein *washout*, which are also present at this locus, are capable of playing a role in wing patterning in *Bicyclus anynana*. Also, Nadeau et al 2016 presents potential evidence for splice variation in *cortex* between different races, but only examined this in whole hindwings. I therefore performed a spatiotemporal transcriptomic analysis on larval and pupal hindwings of individuals from either side of the hybrid zone, with the aim of determining which transcripts are associated with the yellow bar in these two species.

II: MATERIALS AND METHODS

Tissue sampling and dissection

Heliconius melpomene rosina and *Heliconius erato demophoon* were collected from stocks maintained at the Smithsonian Tropical Research Institute in Gamboa, Panama. In addition, hybrid populations of *H. m. melpomene x rosina* and *H. e. hydara x demophoon* were collected around Puerto Lara, Darien Province, Panama. The collected individuals contained a mixture of individuals with $Yb^{+/-}$ and $Yb^{-/-}$ phenotypes, which were returned to the facilities in Gamboa, crossed and selected for the $Yb^{-/-}$ phenotype over one generation to obtain homozygous stocks.

Adults were provided with an artificial diet of pollen/glucose solution supplemented with flowers of *Psiguria*, *Lantana* and/or *Psychotria alata* according to availability. Females were provided with *Passiflora* plants for egg laying (*P. menispermifolia* for *H. melpomene*, *P. biflora* for *H. erato*). Eggs were collected daily, and caterpillars reared on fresh shoots of *P. williamsi* (*melpomene*) or *P. biflora* (*erato*) until late 5th (final) instar, when they were separated into individual pots in a temperature-monitored room, and closely observed for the purpose of accurate developmental staging.

Pre-pupation larvae were identified for dissection. Late 5th instar larvae undergo colour changes from white to purple on the last larval day, followed by an additional change to pink-orange in the hours before pupation. Additionally, several behavioural changes accompany the pre-pupation period; the larvae stop eating and clear their digestive tract, then undertake a period of rapid locomotion and wandering until they find an appropriate perch for pupation - preferably the underside of a leaf or a sturdy twig - at which point they settle in place and produce a strong attachment from their silk glands. Gradually, over a period of 30-120 minutes, they suspend themselves from their perch in a J shape and then pupate. Larvae at the pre-J stage were dissected in cold PBS and the wing discs removed. Whole larval wing discs were transferred into RNAlater and kept on liquid nitrogen in Gamboa, then transported to the UK on dry ice, and transferred to -80°C on arrival in Cambridge.

Pupae were allowed to develop until 36h (+/- 1.5h), or to 60h (+/- 1.5h) post pupation. These time points are referred to as Day 1 and Day 2 throughout. In the hours immediately post-pupation, (Day 0), the pupal carapace is soft and the membranous structures of the pupa are thin, weak, transparent and sticky, hence the effective dissection of unfixed, intact pupal structures is very challenging at the earliest pupal time points.

Pupae were dissected in cold PBS. Wings were removed from the pupa and cleared of peripodial membrane. The hindwings were then cut with microdissection scissors into 2 sections (anterior and posterior), as indicated in figure S3.1. The lacunae (developing veins) were used as landmarks for dissection. As with larval tissue, pupal wing sections were transferred into RNAlater and kept on liquid nitrogen in Gamboa, then transferred to -80°C upon arrival in Cambridge.

RNA extraction and sequencing

RNA extraction was carried out using a standard hybrid protocol. Briefly, wing tissue sections were transferred into Trizol and disassociated using stainless steel beads in a tissue lyser. Chloroform phase extraction was performed, followed by purification with the Qiagen RNA extraction kit. RNA was eluted into distilled water and treated with DNase, then quantified and stored at -80° C. Left and right wings and wing sections were pooled, to increase the yield of RNA extraction.

cDNA synthesis, library preparation and sequencing were carried out by BGI. Samples were sequenced at either 75 PE on Illumina HiSeq 3000 or at 150 PE on Illumina HiSeq 4000.

Mapping and quantification

Reads were aligned with Hisat2 aligner to the genome of the respective species (Kim et al., 2015). The highest percentage of unique mappings was achieved using default parameters (Figure 3.4). Alignments were then quantified using GFF annotations of each genome with HTSeqCount, union mode (Anders et al., 2015). Genomes and annotations are publicly available at www.lepbase.org (Challis et al., 2016).

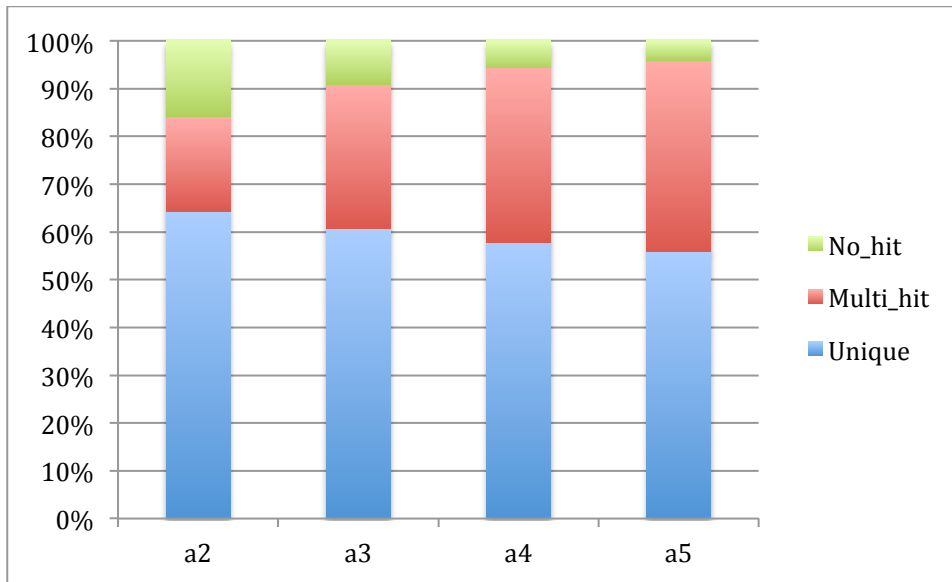


Figure 3.4: Mapping scores to Hera1, varying the score_min parameter in Hisat2, a2 being default. Increasing score_min increases the overall % of reads aligning but decreases the % of reads that have unique alignments: DESeq2 will only count uniquely-mapped reads.

Data analysis

Statistical analysis of counts was carried out using the R package DESeq2 using the following generalized linear model (GLM)

$$\sim \textit{individual} + \textit{compartment}*\textit{race}$$

(Compartments: Anterior Hindwing (HA), Posterior Hindwing (HPo)). Contrasts were then extracted for comparison of race, compartment, and race given the effect of compartment, alternating the race used as the base level. See Figure 3.5 for a graphical depiction of the contrasts.

Gene orthologs between *H. melpomene* and *H. erato* were identified by LepBase using OrthoFinder (LepBase v4, 2017). I verified the LepBase homologs by running BLASTp reciprocally between the *H. melpomene* and *H. erato* gene sets. Gene trees built using InterPro were produced by LepBase.

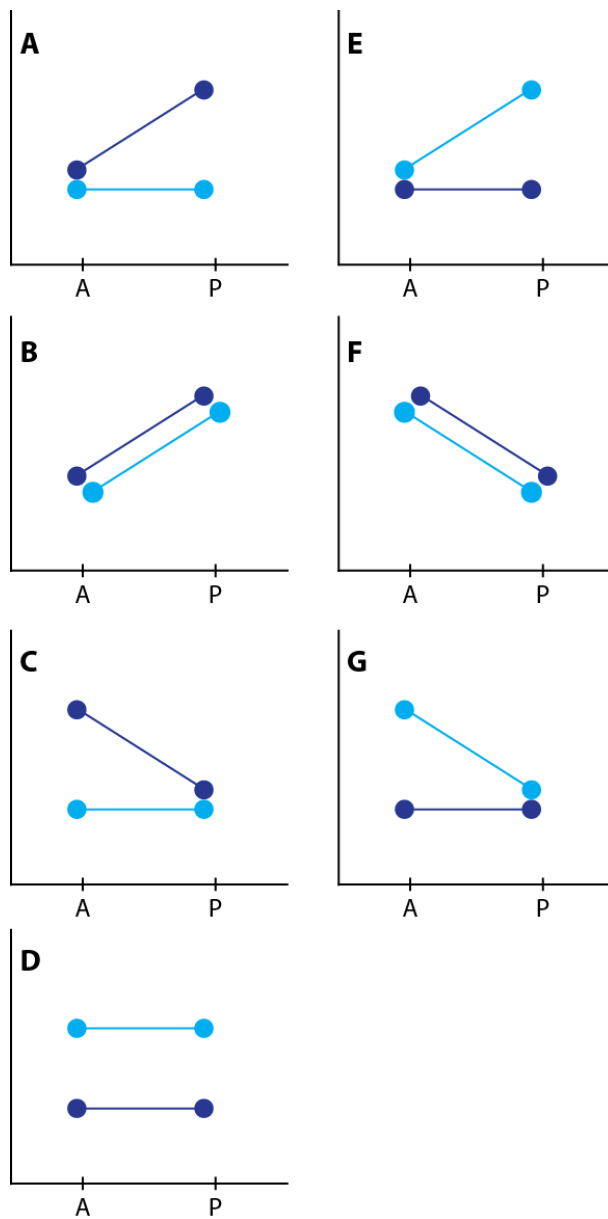


Figure 3.5: Depiction of contrasts. Dark blue represents the yellow races, *H. m. rosina* and *H. e. demophoon* from Panama. Light blue represents the black races, *H. m. melpomene* and *H. e. hydara* from Colombia.

Table 3.1; Gene codes and IDs of genes at the *cortex* locus in *H. melpomene* and *H. erato*.

Hera1 Gene ID	gene name / description	Hmel2 Gene ID
evm.model.Herato1505.61	Acylpeptide	HMEL000003
evm.model.Herato1505.62	Trehalase-1B	HMEL000004
evm.model.Herato1505.63	Trehalase-1A	HMEL000006
evm.model.Herato1505.64	B9 protein	HMEL000007
evm.model.Herato1505.65	HM0008 nov	HMEL000008
evm.model.Herato1505.66	WD40 repeat d 85	HMEL000010
evm.model.Herato1505.67		
evm.model.Herato1505.68	CG2519	HMEL000012
evm.model.Herato1505.69	Unkempt	HMEL000013
evm.model.Herato1505.70	Histone H3	HMEL000014
evm.model.Herato1505.71		
evm.model.Herato1505.72	HM00015 nov	HMEL000015
evm.model.Herato1505.73	HM00016 nov	HMEL000016
evm.model.Herato1505.74	RecQ helicase	HMEL000017
evm.model.Herato1505.75		
evm.model.Herato1505.76	HM00018 nov	HMEL000018
(1505.35, translocation)	BmSUC2	HMEL000019
evm.model.Herato1505.77	CG5796	HMEL000020
evm.model.Herato1505.78	HM00021 nov	HMEL000021
evm.model.Herato1505.79	Enoyl-coA-hydratase	HMEL000022
evm.model.Herato1505.80		
evm.model.Herato1505.81	ATP-bp	HMEL002023
evm.model.Herato1505.82		
evm.model.Herato1505.83	HMEL002023	
evm.model.Herato1505.84	HM00024 nov	HM00024 nov
		g8186
evm.model.Herato1505.85	cortex	HMEL000025
evm.model.Herato1505.86	parn	HMEL000026
evm.model.Herato1505.87	CG31320	HMEL000027
evm.model.Herato1505.88	Manf	HMEL000028
evm.model.Herato1505.89	CG4692	HMEL000029
evm.model.Herato1505.90	PsmD4	HMEL000030
	HM000031 nov	HMEL000031
evm.model.Herato1505.91	Rnz	HMEL000032
evm.model.Herato1505.92	Lmtk1	HMEL000033
evm.model.Herato1505.93	Wdr13	g8202
evm.model.Herato1505.94	Dome1	HMEL013472
evm.model.Herato1505.95	Washout	HMEL000036
evm.model.Herato1505.96	Dome2	HMEL000037
evm.model.Herato1505.97	lethal 2 k05819	g8206

evm.model.Herato1505.98	MKK3	HMEL000039
evm.model.Herato1505.99	ERCC-6	HMEL000040
evm.model.Herato1505.100	Penguin	HMEL000041
evm.model.Herato1505.101	Thymidilate kinase	HMEL000042
evm.model.Herato1505.102	Cas activated DNase	HMEL000043
evm.model.Herato1505.103	Ribosome biosynth r	HMEL000044
evm.model.Herato1505.104	CG12659	HMEL000045
evm.model.Herato1505.105	CG33505	HMEL000046
evm.model.Herato1505.106	Sr	HMEL000047
evm.model.Herato1505.107	Hm00048 nov	HMEL000048
evm.model.Herato1505.108	Hm00049 nov	HMEL000049
evm.model.Herato1505.109		HMEL000050
	Papillin	HMEL000051
evm.model.Herato1505.110	HM000052 nov	HMEL000052

III: RESULTS

Distribution and clustering of data

Broadly, poor sample clustering was observed between each replicate (three per pattern form per species). With the exception of day 1 *H melpomene*, samples did not cluster by compartment, but they did not cluster by individual either (Figure 3.6).

Differential expression analysis

When whole larval hindwings were compared, 785 genes were detected as differentially expressed between *H. melpomene* yellow and black larval hindwings, compared to just 79 between *H. erato* yellow and black. Of these, 16 had clearly definable homologs that were differentially expressed in both species (Table 3.5). *cortex* was differentially expressed in both species, and has the highest adjusted-p value of any gene in *H. melpomene* larvae. Five of these genes had no identified homologs in *Drosophila* or on InterPro, while another 5 were linked to muscle development.

Pupal wings were compared both between phenotype and between wing compartment. In *H. melpomene*, 3550 unique genes were differentially expressed at day 1, and 232 at day 2, while in *H. erato*, 2043 unique genes were differentially expressed at day 1 and 392 at day 2. Of these genes, 1151 identifiable orthologues between the species were differentially expressed at day 1, but just 13 genes were differentially expressed in both species at day 2 (Table 3.6). Genes that were more highly expressed in the black pattern forms at day 1 were enriched for GO annotations linked to mitochondrial respiration GO:0042775, as well as cell cycle and cell division GO:0051297, GO:0007049, GO:0022402 whereas enriched terms in the yellow form include epithelial development GO:0048856, cell polarity GO:0007163, eye development GO:0001654, anterior-posterior axis specification GO:0009952, actin filament-based processes GO:0030029 and wing disc development GO:0007476 (Supplementary Table S3.1).

<i>H.melpomene</i>							
Contrast	A	B	C	D	E	F	G
Larvae	+622-163						
D1	+10, -7	+2271, -1153	+15, -17	+2, -2	+50, -65	+14, -13	+8, -21
D2	+2, -0	+23, -272	+5, -60	0, 0	+124, -52	0, 0	+0 -0
<i>H.erato</i>							
Contrast	A	B	C	D	E	F	G
Larvae	+52 -27						
D1	+1, -4	0, 0	+505-212	0, 0	+580, -1054	+9, -35	+7, -181
D2	+10, -22	+33, -2	+39, -19	+1, -7	+20, -166	+288, -9	+10, -309
Summary							
	Total (<i>H.m</i>)	Total (<i>H.e</i>)	Homologs	homs as % of	homs as % of		
Larvae	785	79	16	2.04%	20.3%		
D1	3550	2043	1151	32.4%	56.3%		
D2	232	392	13	5.2%	3.06%		

Table 3.2: Differentially expressed genes, giving the number of genes up- and down-regulated for each contrast at each stage in each species. Table 1 lists the number of differentially expressed genes, with additional testing of contrasts within the GLM to detect genes DE between compartment given the effect of race, and genes DE between race given the effect of compartment – as illustrated in figure 5. Under “summary”, the total number of unique genes DE in each species is listed, as well as the % of these genes which are DE in both species. A-G correspond to the differential expression in Figure 3.5

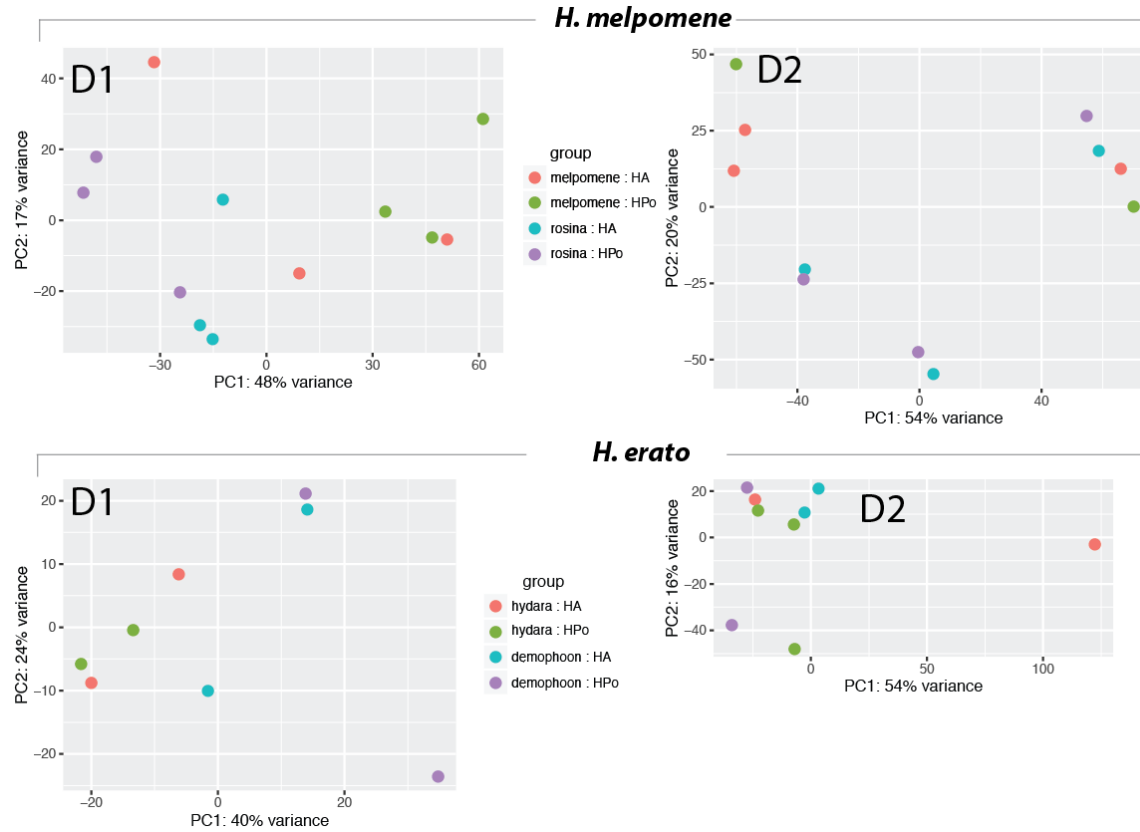


Figure 3.6: Principal component analysis of samples. HA = Anterior hindwing, HPo = posterior hindwing, Day one on left, day 2 on right.

Differential expression at the yellow locus

Expression of genes in the yellow locus broadly supports a role for the previously identified candidate genes *cortex*, *domeless* and *washout*. Figure 3.7 and 3.8 show $-\log$ adjusted p-values (*padj*) for all genes included in the yellow locus by Nadeau et al (2016), as well as the \log_2 FoldChange and figure 3.9 shows transcript abundance for the genes *cortex*, *dome1*, *dome2* and *washout*.

In larvae, *cortex* is differentially expressed in both species. In *H. melpomene*, the novel gene Hmel_000031 is also differentially expressed (Figure 3.7, top). This gene has no homolog in *H. erato*, or with the InterProScan database. In *H. erato*, *cortex* has the lowest adjusted p-value of any differentially expressed transcript in the whole genome (Figure 3.8, top). Notably, *cortex* is more highly expressed in the black form in *H. melpomene* and in the yellow form in *H. erato* (Figure 3.9). *cortex* is the earliest gene to be differentially expressed in both species at this locus.

In day 1 pupae, *washout* is differentially expressed in both species. In *H. melpomene*, many genes at this locus are differentially expressed in contrasts C and F (Figure 3.7, middle). *cortex* has the lowest *padj* score in both of these contrasts. Of the other differentially expressed genes, 5 are also differentially expressed in *H. erato* – *unkempt* (an mRNA-binding ubiquitin ligase), *Histone-H3*, *Hmel000015_nov*, *CG4692* and *CG12659* (the last 4 have no annotated function) (Figure 3.8, middle). In *H. erato*, *wash* is the only gene differentially expressed in contrast C at this locus. Similar to the larvae, *wash* is differentially expressed in the opposite direction between species, in this case more highly expressed in the yellow form in *H. melpomene* and in the black form in *H. erato* – the inverse of *cortex* in larvae (Figure 3.9).

At pupal day 2, no genes were differentially expressed in *H. erato*, and the gene *Trehalase-1B* was differentially expressed in *H. melpomene*.

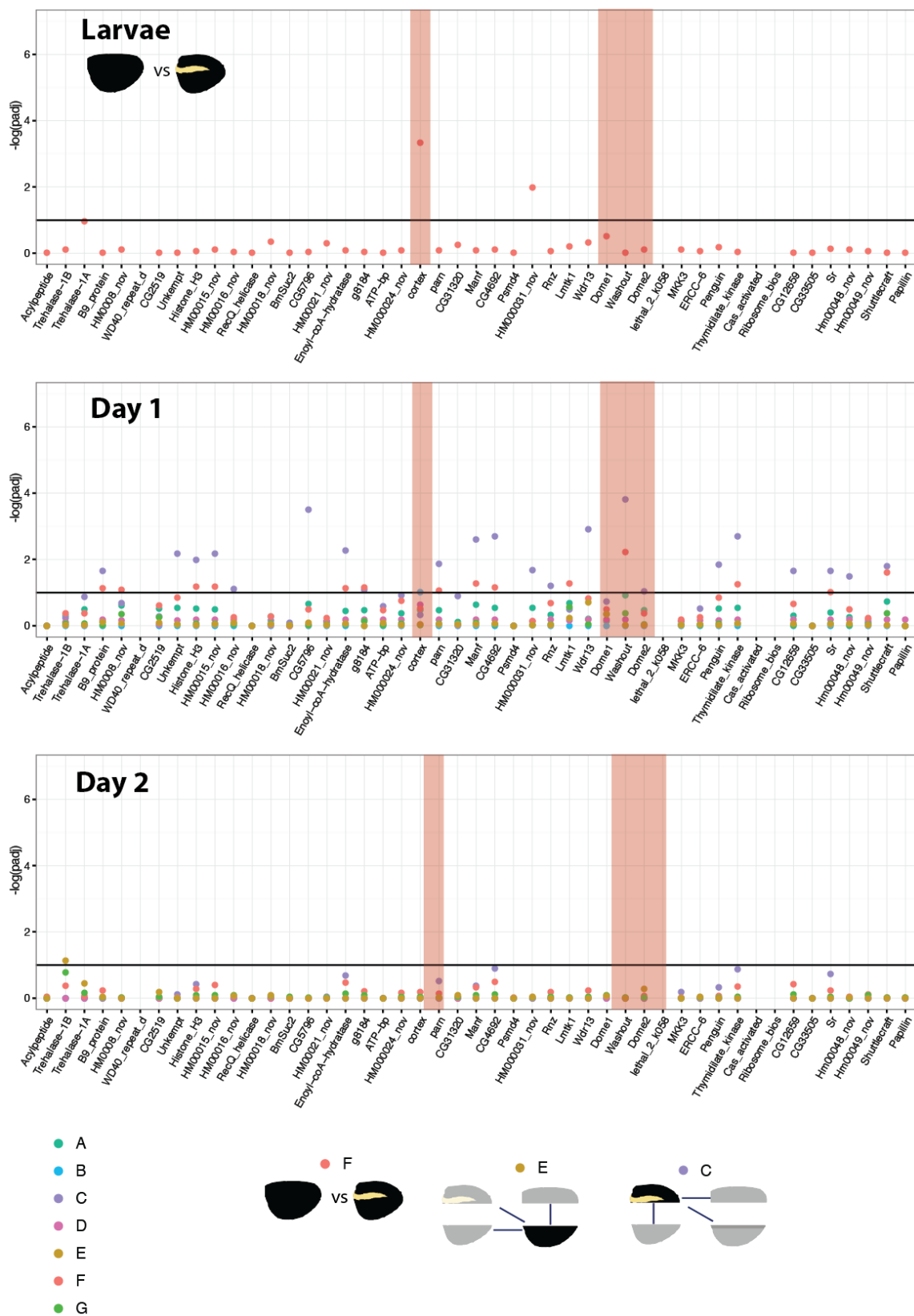


Figure 3.7: Differential expression across the *cortex* locus in *H. melpomene*, shown as the negative log of the adjusted p value ($-\log(p_{adj})$). Top: larvae, middle: day 1 pupae, bottom, day 2 pupae. See Table 2.2 for gene IDs and homology with *H. erato*. Red bars highlight the genes *cortex*, *dome* and *wash*. The horizontal line indicates the cutoff for significance, at $p_{adj}=0.1$. Colours are used for each of the contrasts, depicted in Figure 3.5. In this analysis, genes were differentially expressed in contrast E, C and F (depictions of these contrasts are provided). F is the difference between races, E gives genes differentially regulated in black posterior compartment, and C gives genes differentially regulated in yellow anterior compartment (these contrasts are depicted in cartoon form).

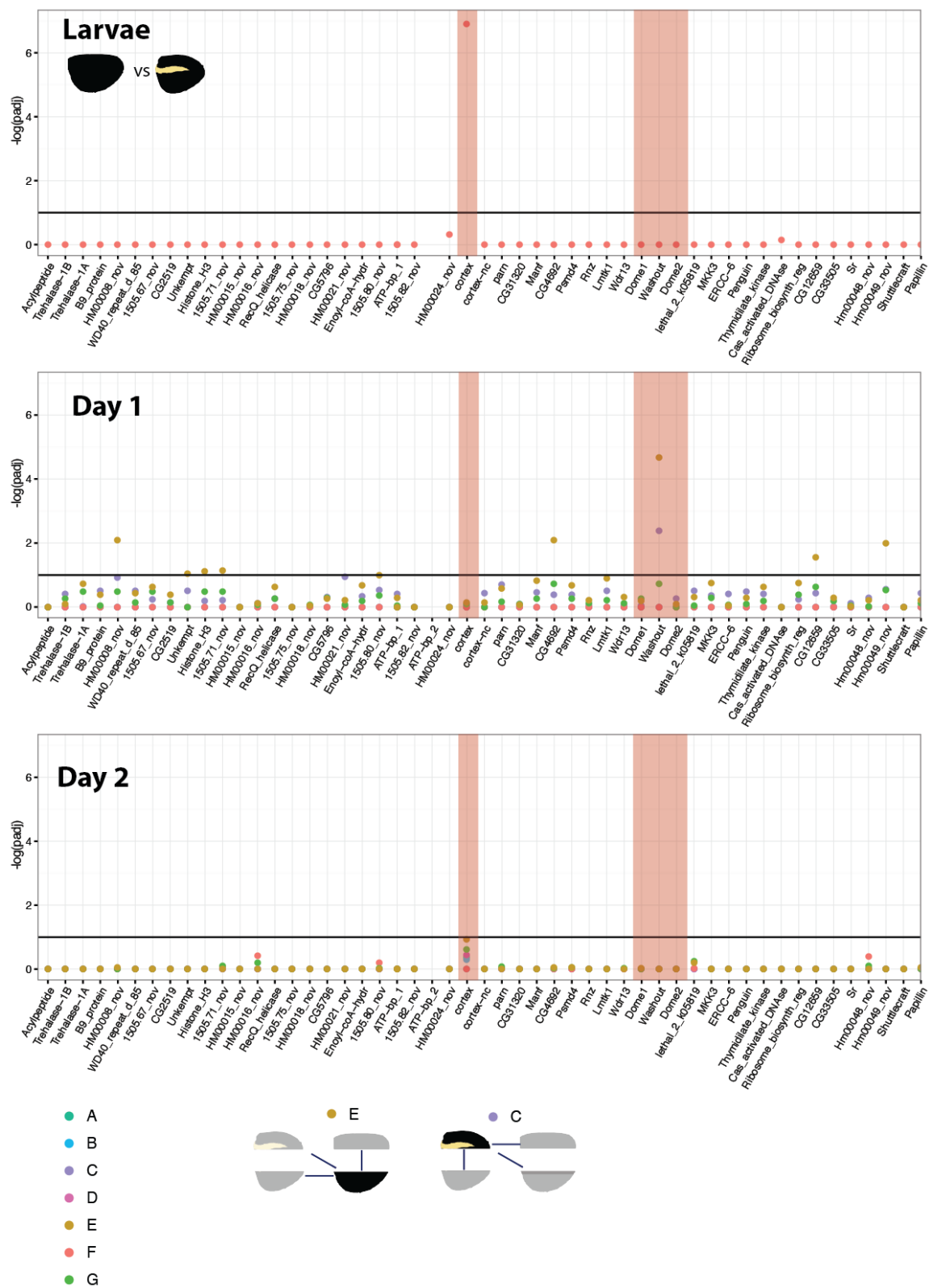


Figure 3.8: Differential expression across the *cortex* locus in *H. erato*, shown as the negative log of the adjusted p value ($-\log(p_{adj})$). Top: larvae, middle: day 1 pupae, bottom, day 2 pupae. See Table 2.2 for gene IDs and homology with *H. melpomene*. The red shading highlights the genes *cortex*, *dome* and *wash*. The horizontal line indicates the cutoff for significance, at $p_{adj}=0.1$. Colours are used for each of the contrasts, depicted in Figure 3.5. In this analysis, genes were differentially expressed in contrast E and C (depictions of these contrasts are provided). E gives genes differentially regulated in black posterior compartment, and C gives genes differentially regulated in yellow anterior compartment.

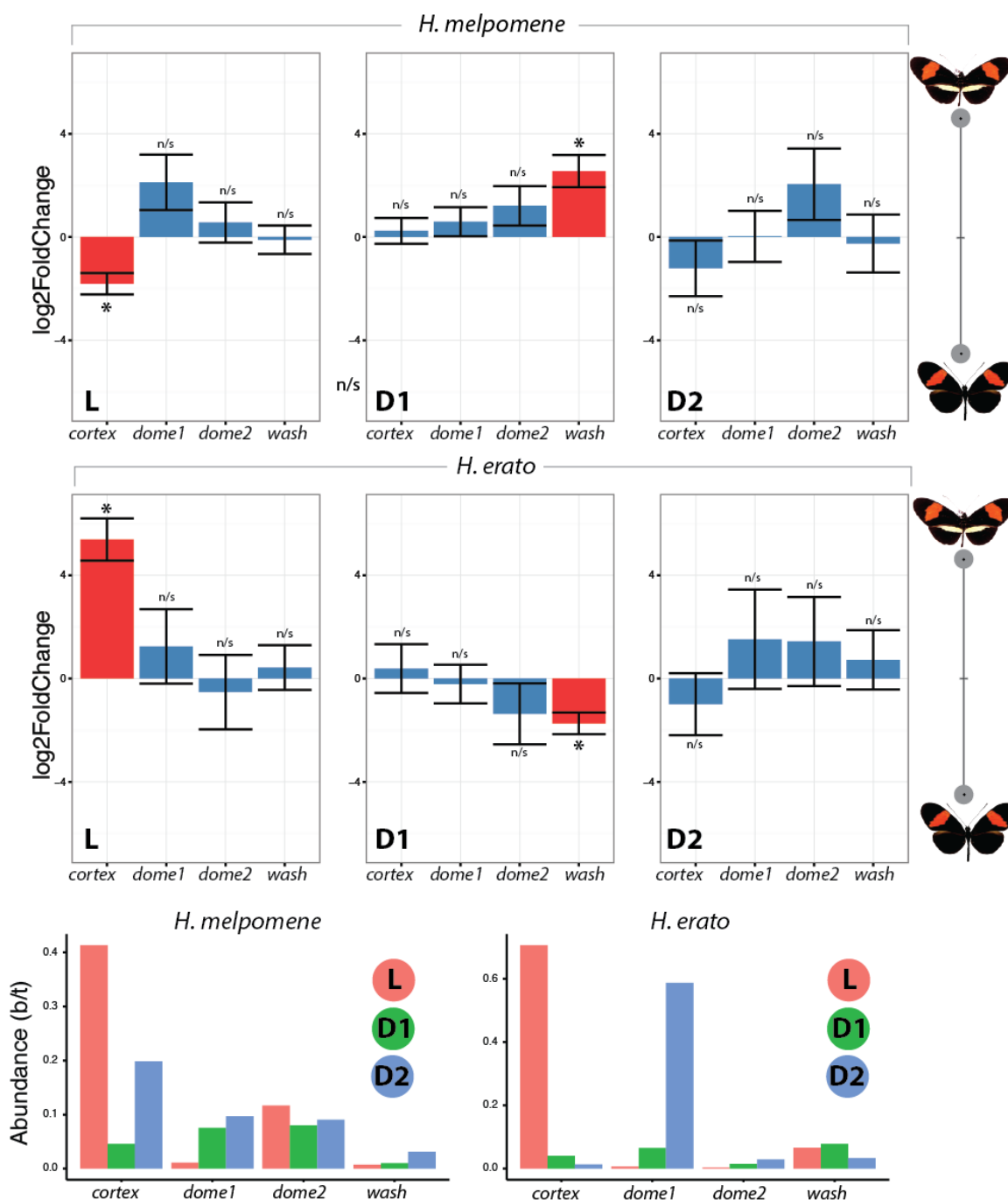


Figure 3.9: Top: Log2FoldChange in contrast C for the genes *cortex*, *dome1*, *dome2* and *wash* in *H. melpomene* and *H. erato*. below, relative abundance of each gene changes through time – *cortex* expression decreases from larval to pupal stages, *dome1* expression increases, whereas *dome2* and *wash* stay relatively constant at all three stages.

The annotation of the *dome*, *dome2* and *wash* genes is especially complex, with overlapping reading frames, so I manually examined the annotation and the raw alignment of reads in IGV for both the Hmel2 and Hera1 annotations (Figure 3.10). The annotation of the 3-prime untranslated region (3'UTR) for *wash* overlaps with the coding gene annotation for *dome2* in both species. It was noted that this includes a number of reads which appear to splice over the coding region of *dome2*. It is not possible to unambiguously assign reads that map to this overlapping portion of the annotation for either gene. Strand-specificity cannot be used to assign reads to the correct transcript, as both genes are coded in the same 5'-3' orientation. Indeed, HTSeq-count calls any reads mapping entirely within overlapping annotations as 'ambiguous', and does not assign them to the final count for either gene. In order to determine if this affected the analysis, I individually dropped each gene in turn from the GFF annotation and re-ran the analysis. When *wash* was dropped, *dome2* was not differentially expressed, and when *dome2* was dropped, *wash* continued to be differentially expressed, implying the reads mapping to the annotation overlap are not driving the fold-change in expression at the *wash-dome2* feature.

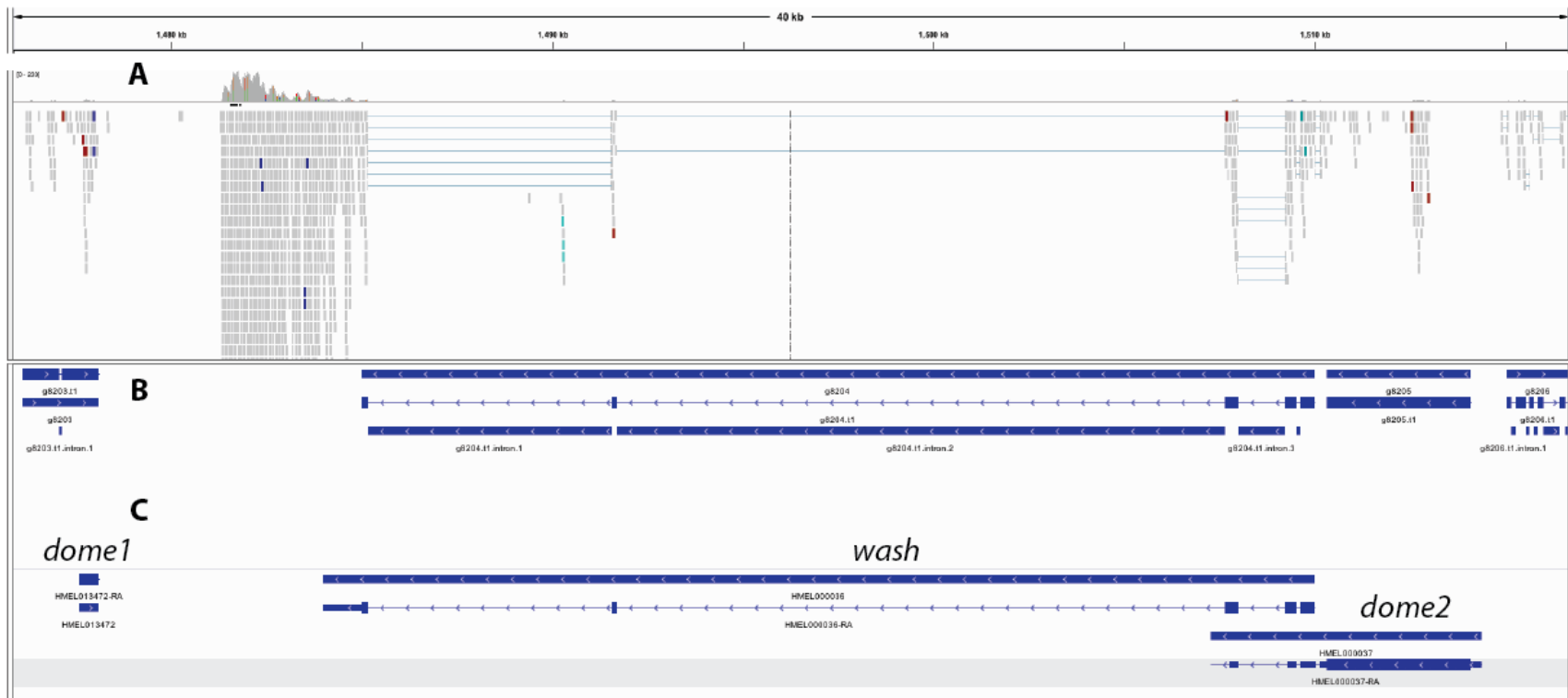


Figure 3.10: IGV raw mapping reads and annotation tracks. A: raw reads from yellow band *H. melpomene* mapped to the *Hmel2* genome by Hisat2. Grey blocks indicate reads, connecting lines indicate spliced read alignments.. B: Hera2.gff from LepBase. C: Hmel2.gff. Exons are indicated by the thick blue lines, interspersed with introns indicated by thin blue lines. Reads map to all three genes, and clear splicing events can be seen between exons. Reads from the 3' end of *wash* flow into the coding portion of *dome2* with no discernible break. A large pile of reads sits immediately 5' to the annotation for the gene *wash*

Additionally, the presence of large non-coding sequences was detected in both *H. melpomene* and *H. erato*, adjacent to the *wash* transcript (Figure 3.10 between *dome1* and *wash*). This was manually annotated as Hera.evm.model.1505-A in Hera1 and Hmel000036A in Hmel2, and covers ~3 kb of sequence between *dome1* and *wash*. This element had high coverage and clear start- and end-points, but contained no identifiable coding sequence, and had no splice sites, implying it is not a coding gene, and not the consequence of genomic contamination. The sequence was compared against itself using BLAST, which indicated no presence of repetitive or low-complexity sequence. BLAST comparisons against the whole Hera1 genome, against all *H. melpomene* sequence on NCBI, and against a *Heliconius* TE library (Lavoie et al., 2013) similarly indicated that this is unique sequence and not repetitive, and is conserved between *H. melpomene* and *H. erato*. The element was not differentially expressed. Another such element was located adjacent to the gene *optix*. It is likely that both elements are polyadenylated long non-coding RNAs (lncRNAs).

The duplication history of these genes was investigated using the gene trees available on LepBase for *dome1* and 2 (Figure 3.11). The duplication of the ancestral *domeless* gene occurred in the common ancestor of *H. erato* and *H. melpomene*, but tandem duplications of *domeless* have occurred in several other Lepidoptera including *O. brumata*, *M. sexta*, *B. anynana*, *Papilio xuthus* and *Plutella xylostella*. Alignment of the proteins indicates that in both *H. erato* and *H. melpomene*, *dome1* is truncated, maintaining only the N-terminal half of the gene. In the *H. e. lattivita* reference annotation, *dome1* appears to have further truncations at the N-terminal end.

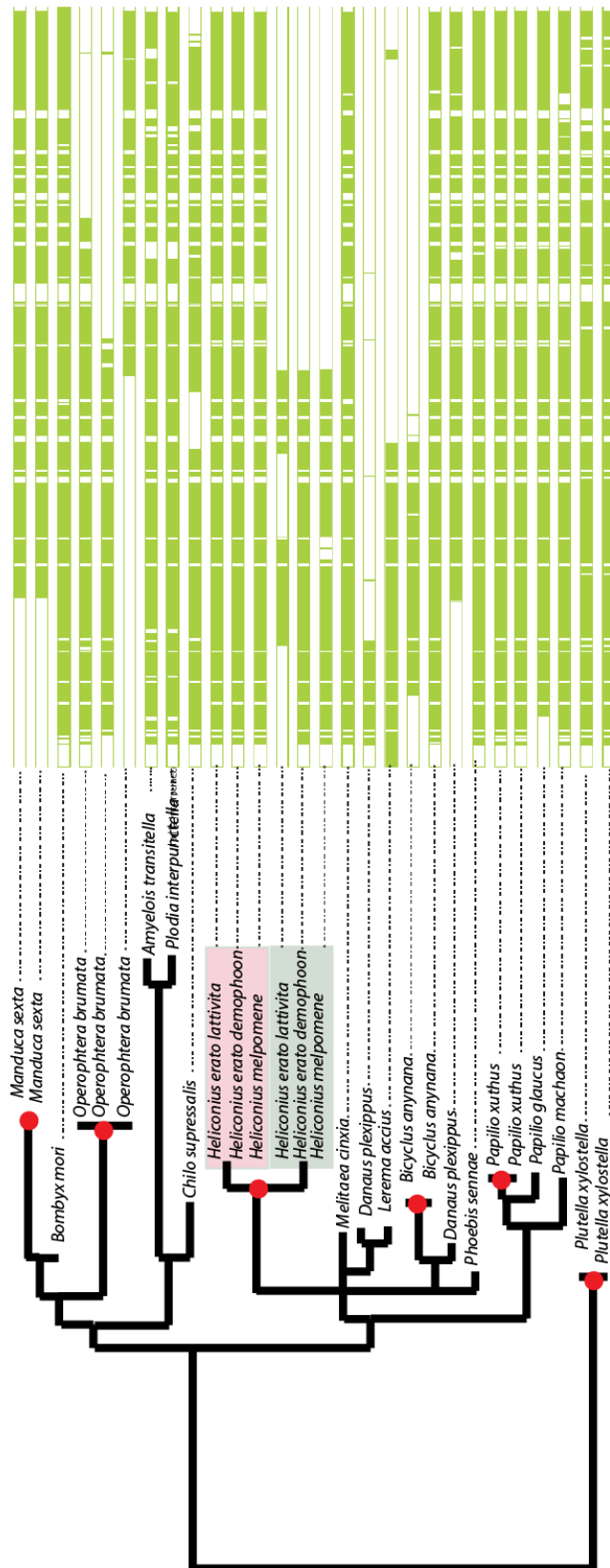


Figure 3.11: LEFT; Gene tree including *dome1* (highlighted in green) and *dome2* (highlighted in pink), retrieved from LepBase. Duplication nodes are highlighted in red. Repeated duplications of *dome* have occurred in multiple lineages. RIGHT: cartoon of the amino acid alignment of the *dome* genes throughout the Lepidoptera. Note the recurrence of C-terminal truncations.

Splice variation of the gene cortex

Previous work has suggested that splice variants might play a role in wing pattern specification during wing development in *H. melpomene* (Nadeau et al., 2016). I compared the mapping of counts and spliced reads to each exon of *cortex* between samples in both species (Figure 3.12 for *H. melpomene* and Figure 3.13 for *H. erato*). If different splice variants were present in the different pattern forms, or in different wing compartments, one would expect the proportion of reads mapping to individual exons to be different, and to fall into two clear groups. Additionally, one would expect to find splice junctions between exons that are present in one group but not another group.

There is evidence for multiple splice variants in both *H. erato* and *H. melpomene*, notably in *H. erato* larvae where reads mapped to exon 2 spliced to either exon 3 or 4, and reads mapped to exon 4 spliced to either 5 or 6. This supports the observation of splice variation in *cortex* described in *H. melpomene* by Nadeau et al, and illustrates that splice variation of *cortex* is also present in *H. erato*.

Notable differences are present in both exon expression and splice junction usage between time points. In larvae, a higher proportion of reads map to coding exons, but E1 is skipped. Later time points have a higher proportion of reads mapping to non-coding exons, accompanied by lower overall expression, including the use of non-coding exons and splice junctions that are not present in the larvae.

However, no consistent difference was observed either between pattern form, or between wing compartment, in either species at any time point. Within each stage, the use of splice junctions did not vary between individuals, and the proportions of reads mapping to exons were either observed to be approximately the same for all individuals, (as with *H. melpomene* larvae) or with high variability which did not group individuals in any meaningful way (eg *H. erato* Day 2).

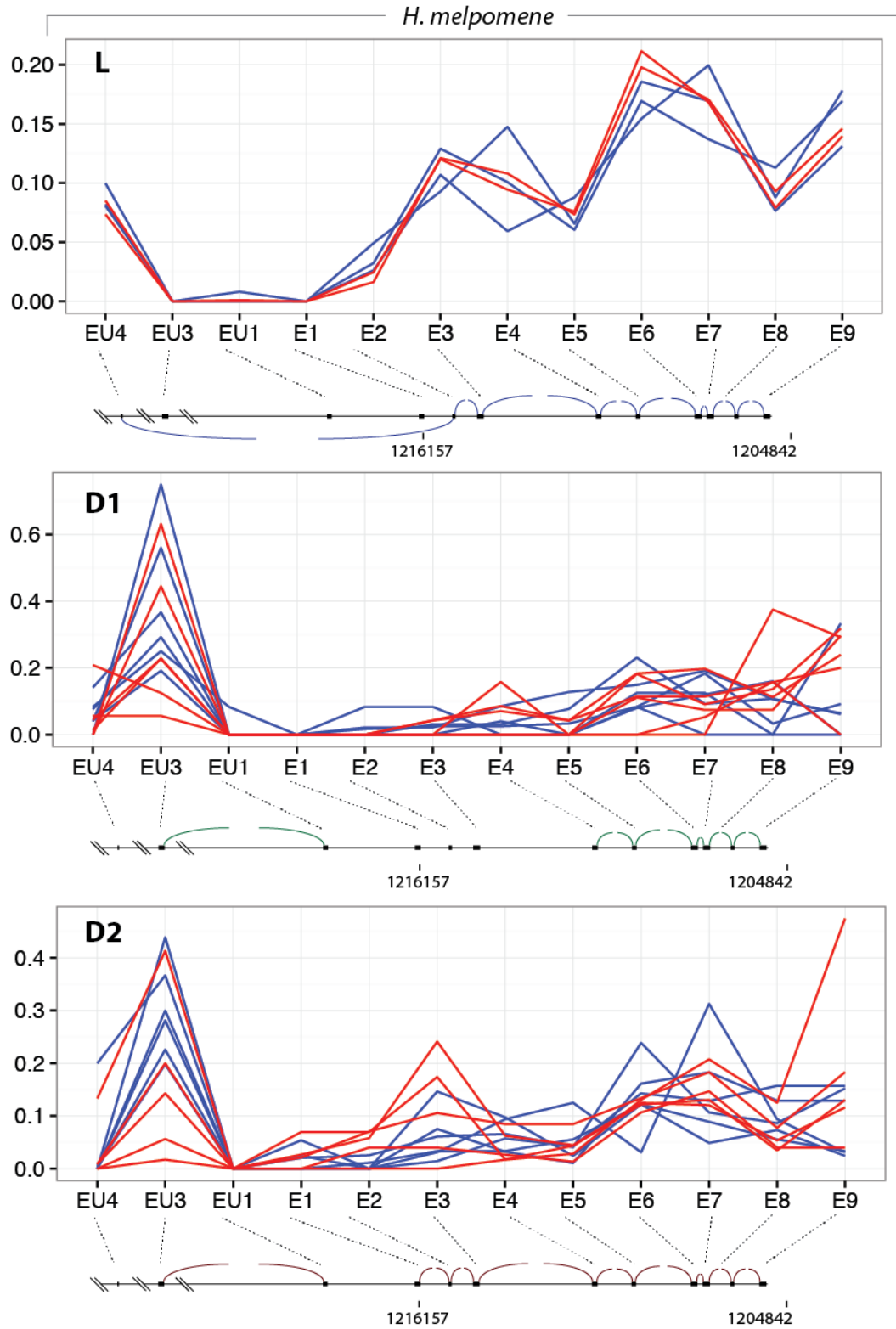


Figure 3.12: Splicing of the gene *cortex* in *H. melpomene*. Blue lines are the black form, red lines are the yellow form. “EU” exons are non-coding exons. “E” exons are

coding exons. The Y-axis is the proportion of reads mapping to each exon. No clear groups of splice variants are apparent.

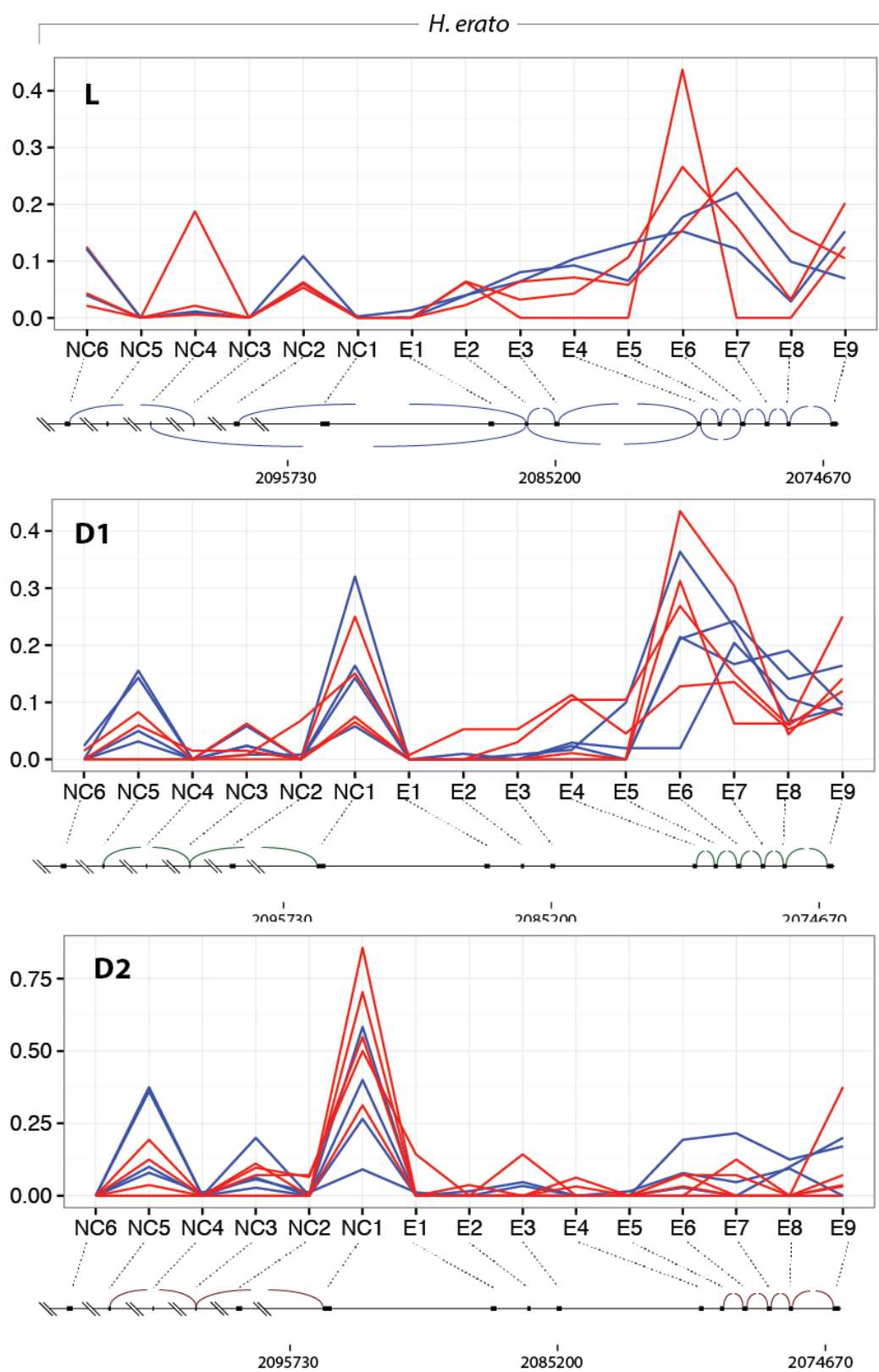


Figure 3.13: Splicing of the gene *Cortex* in *H. erato*. Blue lines are the black form, red lines are the yellow form. No clear groups of splice variants are apparent. “NC”

exons are non-coding, “E” exons are coding. The Y-axis is the proportion of reads mapping to each exon. No clear groups of splice variants are apparent.

Differential expression of Transcription Factors

Particular attention was given to differential expression of transcription factors as they are good candidate genes for controlling expression of the developmental and metabolic networks required to produce different scale cell types.

No TFs were differentially expressed in *H. erato* larval hindwings, whereas in *H. melpomene* the TFs *org-1*, *pnr*, *Rel*, *Sox21a*, *aop* and *Taf7* were differentially expressed between the two races.

At day 1, 53 TFs were differentially expressed in *H. erato* and 72 were differentially expressed in *H. melpomene*. 26 of these TFs were differentially expressed in both species (Table 3.3). These were mainly differentially expressed in contrast F, in which genes differentially expressed between compartment given the effect of race (i.e., genes with the same anterior-posterior expression profile in both races). The genes *FoxK_1*, *Alh* and *mirr* have known roles in antero-posterior patterning. Additional genes are differentially regulated in contrast C, (different in the anterior compartment of the yellow form), or A (different in the posterior compartment of the yellow form).

At day 2, no TFs were differentially expressed in both species, but 5 TFs were differentially expressed in *H. erato*, (*Homothorax*, *TfAP-2*, *cubitus interruptus*, *dumpy* and *jim lovell*) and three in *H. melpomene* (*Hr38*, *Sp1* and *enhancer of yellow-2*) (Table 3.3, lower).

Table 3.3: Transcription factors differentially expressed at Day 1

Day 1; Transcription factors differentially expressed in both species			
Code	Gene	Description	Contrast
NFAT		negative regulator of Ras signal transduction	C, F
p53		apoptosis, cell death, DNA repair	F
org-1	optomotor-blind-related-gene-1	involved in the combinatorial activation of somatic muscle lineage-specific targets	C
Rel	Relish	antibacterial response	F
foxo	forkhead box, sub-group O	compound eye morphogenesis, neural development	F
Myc		contributes to cell growth, cell competition and regenerative proliferation	C F
trx	trithorax	axon guidance, eye development, haltere development, histone methylation	F
grh	grainy head	cell shape, epithelial morphogenesis, cuticle depositoin	C A
tgo	tango	muscle development, R7 cell differentiation	F
E2f1	E2F transcription factor 1		C F
TfAP-2	Transcription factor AP-2	mediates Notch signaling in the developing tarsus	C
pdm3a	pou domain motif 3	axon targeting of olfactory neurons	F
pdm3b	pou domain motif 3	axon targeting of olfactory neurons	F
jumu	jumeau	asymmetric protein localization, chromatin modification, dendrite formation and organ (eye, wing and bristle) development	B C
cwo	clockwork orange	circadian regulation of gene expression	F
Nf-YA	Nuclear factor Y-box A	lateral inhibition, R7 cell differentiation, regulation of cell cycle	F
CG7368		predicted TF	C F
FoxK_1	Forkhead box K	direct target of Dpp/BMP signaling during midgut development, and in turn regulates the homeobox gene lab to determine endoderm differentiation	F
phtf		predicted TF	F
Hr96	Hormone receptor-like in 96	binds cholesterol, required for lipid homeostasis	F
CG3328		predicted TF	F
Alh	Alhambra	maintains eve expression in nervous	F

		system, also req for muscle development	
mirr	mirror	anterior-posterior patterning	C
sens	senseless	stimulates expression of proneural gene expression, PNS and R8 cell differentiation	F
dl	dorsal	downstream of Toll pathway, D-V patterning	F
Day 2, <i>H. erato</i>			
Hth	<i>Homothorax</i>		A,B,C,D
TfAP-2			E
ci	<i>cubitus interruptus</i>		G, F, E
dpy	<i>dumpy</i>		G, F, E
lov	<i>jim lovell</i>	gravitaxis, ventral midline development	D, A, G
Day 2, <i>H. melpomene</i>			
Hr38	Hormone receptor-like 38	required for the proper development of the adult cuticle	C,
Sp1		leg morphogenesis	C, F
e(y)2	<i>enhancer of yellow-2</i>		B

Differential expression of Wnt pathway genes

Particular attention was also given to Wnt-pathway component genes, as recent CRISPR-mediated knockouts of *WntA* in multiple species of *Heliconius*, (including *H. erato*, *H. melpomene*, as well as *H. sara*) have indicated that in the absence of WntA, the yellow bar expands towards the anterior edge of the wing, into the overlap region (though it never expands past the posterior boundary of the wild-type pattern). *WntA* is not the target of selection in the Panama yellow band hybrid zone, but is likely a component of the gene regulatory network that shapes hindwing pattern.

In larvae, no Wnt pathway genes are differentially expressed in either species. Figure 3.14 illustrates differential expression of Wnt-pathway genes in both species at day 1, (grouped by generalized role in the pathway – ligand, receptor, planar cell polarity, nuclear, intracellular, and heparin-interacting). 16 genes are differentially expressed in *H. erato* (including *WntA*), and 24 are differentially expressed in *H. melpomene*. Several Wnt pathway genes are differentially expressed in both species (Table 3.4). These genes are mainly differentially expressed in contrast B, in which genes are differentially expressed between compartment given the effect of race (i.e., genes with the same anterior-posterior expression profile in both races). The exceptions are the genes *hyrax* and *pontin*, transcriptional co-activators which are also differentially expressed in contrast C (different in the anterior compartment of the yellow form).

At day 2, no Wnt pathway genes were differentially expressed in *H. erato*, and only one gene was differentially expressed in *H. melpomene*, - *dally*, or *division abnormally delayed*, a core protein of heparan sulfate proteoglycans, which acts as morphogen co-receptor.

Table 3.4

Wnt genes differentially expressed in both species at day 1			
Code	Gene	Description	contrast
<i>bsk</i>	<i>basket</i>	Serine-threonine kinase, phosphorylates <i>jra</i> (JNK pathway) and	B
<i>dsh</i>	<i>dishevelled</i>	interacts with <i>Fz</i> , signal transduction	B
<i>sgl</i>	<i>sugarless</i>	heparan sulfate biosynthetic process	B
<i>sgg</i>	<i>shaggy</i>	GSK3, β -catenin destruction complex component	B
<i>Drl-2</i>	<i>Derailed-2</i>	Wnt Receptor	B
<i>Axin</i>		interacts with <i>Fz</i> , signal transduction	B
<i>mwh</i>	<i>multiple wing hairs</i>		B
<i>dco</i>	<i>discs-overgrown</i>	Ser/Thr protein kinase, planar cell polarity	B
<i>hyrax</i>		recruited by <i>arm</i> and <i>ci</i> , transcriptional coactivator.	B C
<i>wls</i>	<i>Wntless</i>	transmembrane protein required for Wnt ligand secretion	B
<i>pontin</i>		Transcription coactivator	B C

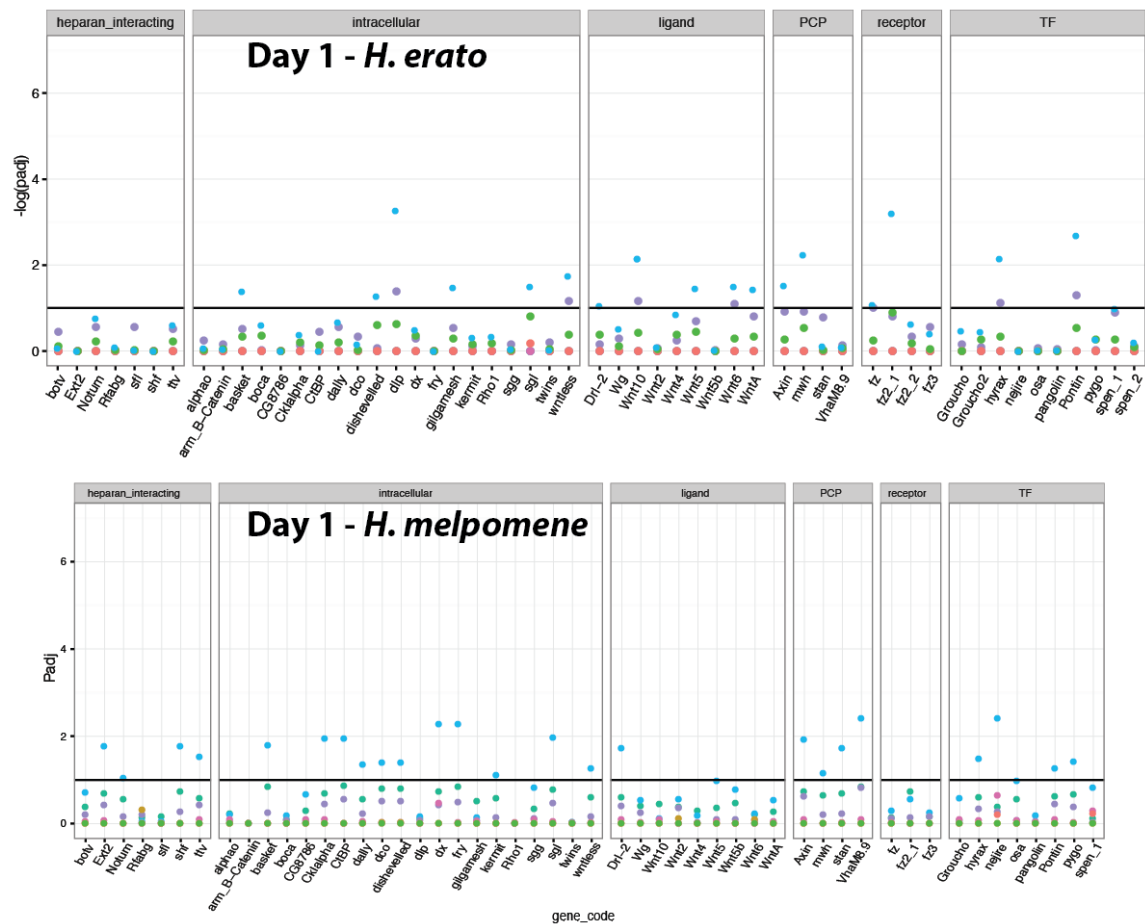


Figure 3.14: Differential expression of Wnt pathway constituents in *H. melpomene* and *H. erato* in day 1 pupae. Upper panel shows *H. melpomene*, lower panel shows *H. erato*. Genes are grouped by their general function in the Wnt pathway.

Table 3.5

Larvae, differentially expressed homologs			
era_gene	melp_gene	alias	notes
Herato1505.85	HMEL000025	<i>cortex</i>	linked to yellow patter elements
Herato1603.22	HMEL006576	???	no predicted homolog
Herato1202.408	HMEL010272	<i>thin</i>	aka abba. Ubiquitin ligase, muscle cell fibre development, myofibril assembly.
Herato1408.66	HMEL011448	???	no predicted homolog
Herato1408.65	HMEL011450	???	no predicted homolog
Herato1605.151	HMEL011803	<i>CG16885</i>	homology to human ZF domain protein
Herato1202.675	HMEL011976	<i>methuselah</i>	GPCR, cell surface receptor, determination of lifespan, starvation response
Herato1805.130	HMEL016193	<i>m-spondin</i>	regulation of myoblast function
Herato0601.83	HMEL016518	???	no predicted homolog
Herato0606.71	HMEL016787	<i>CG43897</i>	homology to sNPF (short neuropeptide-F). Lateral inhibition, mesoderm development
Herato1202.23	HMEL032945	<i>CG18420</i>	proteolysis (predicted)
Herato1301.673	HMEL034755	???	no predicted homolog
Herato0209.59	HMEL036844	<i>CG6142</i>	(homology to glucose dehydrogenases) oxidoreductase activity.
Herato2101.520	HMEL042582	<i>Paramyosin</i>	structural component of muscle - (thick filament).
Herato0419.35	HMEL044001	<i>CG3409</i>	(homology to Monocarboxylate transporters),
Herato0609.73	HMEL045659	<i>Cp65Ec</i>	Homology to 'Bmori-larval cuticular protein'.

Table 3.6

Table 6: Day 2 differentially expressed homologs			
era_gene	melp_gene	alias	notes
Herato0411.101	HMEL014400	<i>disco-r</i>	'related to disconnected', wing expansion, epithelial-mesenchymal transition, wing
Herato0606.333	HMEL045330	<i>Gelsolin</i>	actin filament polymerisation
Herato0801.48	HMEL009162	<i>RBP2a</i>	'Reticulocyte-binding protein 2 homolog a' (homology to heparin binding proteins)
Herato1301.288	HMEL034236	<i>HMEL034236g</i>	*** no homologs ***
Herato1301.293	HMEL006011	<i>Cp65Av</i>	'Cuticular protein 65Av'
Herato1505.34	HMEL035660	<i>LPS-BP</i>	'Hemolymph lipopolysaccharide-binding protein'
Herato1603.54	HMEL012304	<i>Myofilin</i>	interacts with ap, Bx, Chi, ssdp. Chetae development, wing development.
Herato1701.267	HMEL038231	<i>discs-lost</i>	interacts with CycE, imaginal disc morphogenesis, cell proliferation, lateral inhibition, dev. of polarised epithelium.
Herato2101.520	HMEL042582	<i>Paramyosin</i>	structural component of muscle - (thick filament).
Herato0101.746	HMEL011985	<i>aristaless</i>	antenna development, chaeta development, leg development,
Herato1301.707	HMEL020501	<i>HMEL020501g</i>	fibroin-domain containing
Herato0701.562	HMEL037600	<i>HMEL037600g</i>	*** no homologs ***

IV: DISCUSSION

A major goal in evolutionary biology is to link patterns of selection in wild populations to variation in DNA sequence in the genome. The yellow band hybrid zone in eastern Panama has been well studied and is an example of a phenotype under strong natural selection. Previous work has indicated that the gene *cortex* is the primary target of selection controlling this phenotype. Here, I have provided additional evidence from patterns of gene expression that *cortex* is the earliest differentially expressed gene between pattern forms that vary by the presence or absence of yellow pattern elements. Nonetheless, differential expression of additional candidate genes, including *wash*, indicates that the locus may be more complex than has been thought, potentially containing multiple functional genes.

Differential expression and sample clustering

In principle component analysis, samples did not cluster by compartment or race; typically, such clustering is expected in an RNAseq experiment. However, samples also did not cluster by individual, indicating that the lack of clustering may not be related to poor experimental design. Differential expression of *cortex* was detected, recapitulating the result of Nadeau et al (2016), which indicates that despite poor clustering, the experiment does have the statistical power to detect differential expression of relevant genes.

Dynamics of differential expression

At the earliest time point in larvae, only *cortex* and some muscle related genes are differentially expressed in both species. The early differential expression of *cortex* in the absence of the differential expression of other genes supports the hypothesis from Nadeau et al (2016) that it is the primary gene underlying the patterning function of the yellow locus. Many genes are differentially expressed at day 1, and these genes may be downstream effectors of *cortex* which are responsible for differentiation of Type I scales. The differential expression of genes at the *cortex* locus, including *unkempt*, *dome/wash*

and others, subsequent to the expression of *cortex* could either indicate that they are downstream of *cortex*, or alternatively that they are differentially regulated as a secondary target of the *cortex cis*-regulatory elements. By day 2, fewer genes are differentially expressed; this is in part driven by a lower discovery rate from differential expression analysis, but may also reflect the fact that, after an earlier period of cell type differentiation, the scale cells begin to develop in broadly the same way, requiring scale and wing development factors like *aristaless*, *discs-overgrown*, and actin related genes like *Gelsolin*.

The gene *parn*, which is within the F_{st} window from the comparison of *H. e. hydara* vs *demophoon* in Van Belleghem et al (2017), is never differentially expressed, indicating that despite its proximity to regulatory sequences that code for pattern differences, the gene is not likely to participate in wing pattern formation.

One surprising discovery is the difference in the direction of *cortex* and *dome/wash* expression between *H. erato* and *H. melpomene*. I suggest two hypotheses that could account for this. First, it is possible that the two genes have opposing functions in the two species. Derivation of this yellow pattern element is likely independent in the *H. melpomene* and *H. erato* lineages (as indicated by the fact that the yellow bar has evolved twice within the *H. erato* lineage), so this could represent the repurposing of a wing patterning hotspot but via a different molecular mechanism, i.e. the exchange of roles from activator to repressor. Second, it is possible that the genes have highly dynamic expression, and while they are consistently differentially expressed, the exact pattern of expression changes rapidly through time. In this scenario, both developmental heterochrony and relatively small discrepancies in staging between the two species could cause expression in the opposite direction. It has been suggested that *cortex* is involved in the regulation of the cell cycle. Some families of genes have expression profiles that fluctuate with the cell cycle, most prominently the cyclins, and a number of cell cycle linked genes including Nf-YA are also differentially expressed in this analysis, so this could also indicate some role for cell cycle regulation in the differentiation of Type I scales.

It would be possible to test for dynamic expression of *cortex*, *dome* and *wash* using a similar experimental design to the classic clock-and-wavefront experiments performed on Chicken segmentation (Palmeirim et al., 1997). Butterfly larvae are known to be highly tolerant to surgical removal of imaginal discs (for example, see <http://www.biographic.com/posts/sto/lens-of-time-building-a-butterfly-wing>). One wing could be surgically removed from a larva and fixed, then after a fixed period of time the other wing could be removed and fixed. This would allow the observation of changes in expression domain of a highly dynamic gene without the need for highly synchronized or accurate staging.

Confirmation of pattern-linked expression of cortex

Nadeau et al showed that *cortex* was differentially expressed in the forewing between pattern forms of *H. melpomene* (specifically Amazonian *H. m. malleti* which has the dennis-ray pattern and Ecuadorean *H. m. plesseni* both pictured in Figure 1.1), that *cortex* is expressed in larval wings of *H. m. rosina* in a pattern associated with the yellow band, and that there was evidence for differential exon usage between different pattern forms – in particular, that exon 5 was absent in *H. m. melpomene* but present in *H. m. rosina* (which differ only by the presence or absence of the yellow band).

Here, I confirm that *cortex* is also differentially expressed between other pattern forms of *H. melpomene*. Between yellow *H. m. rosina* and black *H. m. melpomene*, *cortex* was more highly expressed in black regions and the black pattern form, which corroborates the previous finding that *cortex* is differentially expressed in *H. melpomene* proximal forewings consistent with forewing pattern elements. My results also corroborate the *in situ* hybridization expression profile obtained for *cortex*, where staining was detected throughout the wing except for the anterior portion that would become the yellow bar. Based on this expression pattern and the differential expression result herein, I would predict that in the black form of *H. melpomene*, the domain of *cortex* expression would extend up through the whole of the larval hindwing.

Unlike the result described by Nadeau et al, I found no evidence for differential use of

exons or splice sites between different pattern forms. This could reflect the fact that Nadeau used whole hindwings rather than dissected hindwings, and also used different time points of pupal development from those used here, although this is not likely as the difference in usage of exon 5 is consistent through all stages between black *H. m. melpomene* and yellow *H. m. rosina* in that analysis. As with the difference in the directionality of expression, the difference of observed result could also be related to highly dynamic changes in abundance of *cortex* transcripts.

Differential splicing of *cortex* was observed between stages, which was not observed by Nadeau. Both *H. melpomene* and *H. erato* have multiple non-coding exons, which are positioned at a substantial distance from the protein-coding portion of the gene. In fact, the first non-coding exon is 160 kb from the 3' end of the gene. The length of non-coding sequence, and its variance between stages, will have consequences for the stability of the transcript, as well as the timing of expression. For example, base-pair elongation of RPol2 is approximately 20 bp/s (Tolic-Norrelykke et al., 2004) (velocity measure from *E. coli*). This means that to transcribe the whole *H. melpomene cortex* locus, at over 165,000 bp, would take 8283 seconds, or 138 minutes. Genes with exceptionally long 1st introns are recorded elsewhere in the arthropods, including the *Drosophila Hox* genes. It has been suggested that long introns may play a role in temporal control of gene expression (Pace et al., 2016). Also, the variants in non-coding exons observed between stage cause significant length-changes between transcripts, which is likely to modify the relative stability at different times. This will also have consequences for temporal control of transcript abundance. Ultimately, this does not clarify the functional mechanism by which *cortex* or (*dome* and *wash*) can affect differentiation of Type-1 scale cells, but it does suggest several mechanisms by which this process may be regulated.

A role for dome and wash

dome is the only described invertebrate receptor for the JAK/STAT signaling pathway (Brown et al., 2001). It includes fibronectin domains, a transmembrane domain, and a cytokine binding domain. In *D. melanogaster*, it is expressed both maternally and embryonically, in tracheal pits, spiracles and the CNS. Mutants of *dome* have embryonic-

lethal effects on segmentation, indicating a role in regulating the segment polarity genes.

washout (Wiskott-Aldrich scar homolog protein), is a WASP family member and effector of Rho-GTPase that interacts with the Arp2/3 complex, thus contributing to actin polymerization (Derivery and Gautreau, 2010, Tomancak et al., 2002, Rodriguez-Mesa et al., 2012). In *D. melanogaster*, it is expressed in all epithelia but is enriched in spiracles and Malphigian tubules. Mutant phenotypes indicate a role for regulation of dorsal-ventral patterning and regulation by segment polarity genes.

dome and *wash* have not previously been identified as candidates for wing patterning in *Heliconius*, however they have been identified as the likely causative genes for the *Bigeye* family of mutants in *Bicyclus anynana* which are a series of laboratory pattern mutants that have been mapped to a genome region very close to *cortex* (Saenko et al., 2010) (Carolina de Silva, Masters Thesis). Based on embryonic and larval expression domains, and on the embryonic phenotypes of homozygous *Bigeye* mutants (which is lethal), it was hypothesized that in *Bicyclus*, one or both of *dome* or *wash* are likely regulators of the Wnt signaling pathway, and of *engrailed*. The identification of larval and pupal expression of *wash* and *dome* in *H. erato* and *H. melpomene* suggests that they could play a similar role in wing patterning, which is supported by the fact that Wnt-pathway genes were found to be differentially expressed in both species, and the transcription factor *invected*, which is functionally redundant with *engrailed*, is differentially expressed in *H. melpomene*. This could provide an indication that these factors are constituent parts of a gene regulatory network that can influence wing pattern. In addition, the gene *basket*, which is part of the Janus Kinase signaling pathway (downstream of *dome*), was also differentially expressed in both species at day 1, further implicating JAK-STAT signaling.

Gene trees of *dome* indicate that the duplication into *dome1* and *dome2* is basal to the *Heliconius* clade. The *dome1* copy has a truncated C-terminal domain, whereas *dome2*, which overlaps with *wash*, contains a full-length homolog of *dome* in other species. Notably, several species included on LepBase have tandem duplicates of *dome*, including *O. brumata*, *M. sexta*, *B. anynana*, *P. xuthus* and *P. xylostella*. However, none of these

duplicates exhibit the C-terminal truncation. This could imply a convergent selective pressure for duplication of *dome* in many species, or it could imply that the duplication is ancestral but that gene conversion has made the gene sequences similar between the two copies.

Overlapping and abutting annotations, as observed at *wash* and *dome*, are commonly found in viruses and prokaryotes, and occasionally in eukaryotes (McLysaght and Guerzoni, 2015). The close apposition of *dome* and *wash* can also be observed in *B. anynana* (Saenko et al., 2010), as well as *D. plexippus*, *L. accius*, *M. cynxia*, *P. glaucus* and *B. mori* (retrieved from LepBase), implying that the synteny of these genes is conserved across Lepidoptera. These genes have obvious mechanisms by which they might interact with scale development, and are good candidates for an ancestrally conserved wing patterning pathway constituents.

In order to determine if *wash*, *dome* or other genes play a functional role in wing pattern formation, it will be necessary to generate CRISPR-mediated mutants. It is likely that complete homozygous mutants of *wash* and *dome* will be embryonic-lethal (based on *D. melanogaster* phenotypes), but much success has been achieved in generating clonal G₀ mutants of lethal genes including *abdominal-A* with the CRISPR system, both in the butterfly *Papilio xuthus* (Li et al., 2015) and in the amphipod crustacean *Parhyale* (Martin et al., 2016).

Input from Wnt and TFs

The Wnt pathway plays a clear role in laying down wing patterns in *Heliconius*. The majority of Wnt pathway genes that were differentially expressed were expressed between anterior and posterior compartment regardless of race, with the exception of *hyrax* and *pontin*, transcriptional co-activators, which could therefore be downstream of *cortex*. In CRISPR-cas9 mediated knockouts of *WntA*, the hindwing yellow bar has been shown to expand at the anterior edge. Thus, it is likely that the anterior edge of the yellow bar is defined by the expression of *WntA* – though no expression profile for this gene has been published in hindwings. *WntA* was shown to be upregulated in the anterior

compartment in *H. erato*, but as the dissection scheme used here does not give finer detail, this cannot be correlated with the anterior boundary of the yellow bar. Many of the factors identified as differentially expressed between anterior and posterior compartment here were also identified in the comparative analysis in chapter 4

A variety of transcription factors are also differentially expressed during pupal development, including a number of factors related to axis specification, eye development, wing development and neural development. Some of these transcription factors may be downstream effectors of *cortex*, and are playing a role in the differentiation of Type I scales. One key example is *e(y)2*, *enhancer of yellow-2*, perhaps involved in differential deposition of melanin.

Manual correction of annotations

Manual checking of the annotations and read alignments at the *cortex* locus indicated the presence of an overlap between two relevant genes, the premature truncation of annotations of some protein coding genes, and the absence of annotation of candidate lncRNAs. This mis-annotation, along with the total absence of some classes of genes (i.e. non-coding genes) highlights how untested transcript annotations require extensive manual validation and optimisation. Errors and mis-annotations are the result of a combination of the use of automated computational annotation, the use of incomplete tissue libraries to train the annotator. Incomplete or truncated annotations have the potential to non-uniformly decrease counts for genes genome-wide, which is likely to reduce the power of any statistical analysis. Updated annotations were noted and will be added to the online version of the annotations for Hera1 and Hmel2 at Lepbase.org. Manual curation of a whole transcriptome annotation is not feasible, but as the genome and annotation are increasingly used by the lepidopteran genomics community, gradual improvements to annotation will hopefully accumulate and create better resources for future analyses.

Candidate lncRNAs were identified. While neither of these elements were found to be differentially expressed between races or compartments in this experiment, it is important

to catalogue the presence of such transcripts both at colour pattern loci and genome wide, as they may play roles in regulation of transcription, translation, splicing, epigenetic modifications, differentiation and development, which may be relevant to the function of these or other loci (Fatica and Bozzoni, 2014). Some efforts have been made to identify and annotate sncRNAs including miRNAs (Surridge et al., 2011) and piRNAs (Pinharanda, unpublished).

Summary

In the considerable body of work on wing pattern development and diversity in *Heliconius*, the *cortex* locus remains the most enigmatic locus. While it is simple to make predictions about the mechanism by which a transcription factor (*optix*) or a signaling ligand (*WntA*) might feed into the differentiation of scale cells, the functional basis of this locus remains obscure. I have provided evidence that *cortex* is the earliest differentially expressed gene at this locus but also that *wash* and *dome* may play a role. Most curiously, I have illustrated that these genes are associated with wing pattern in both co-mimics *H. melpomene* and *H. erato*, but are correlated with the opposite wing pattern. I have also ruled out a role for splice variation in *cortex* as the causative agent of wing pattern variation in this case. While it remains difficult to decode this part of the gene regulatory network of wing patterning in the absence of manipulative experimental data, this work has offered insight into the processes by which the causative agents at this locus are differentially regulated to steer pattern development.

Supplementary Table 3.1 Gene Ontology terms enriched in black

Go term	Description
(GO:0046034)	ATP metabolic process
(GO:0042773)	ATP synthesis coupled electron transport
(GO:0009058)	biosynthetic process
(GO:0007049)	cell cycle
(GO:0022402)	cell cycle process
(GO:0043603)	cellular amide metabolic process
(GO:0044249)	cellular biosynthetic process
(GO:0071840)	cellular component organization or biogenesis
(GO:0034645)	cellular macromolecule biosynthetic process
(GO:0044260)	cellular macromolecule metabolic process
(GO:0044237)	cellular metabolic process
(GO:0044271)	cellular nitrogen compound biosynthetic process
(GO:0034641)	cellular nitrogen compound metabolic process
(GO:0009987)	cellular process
(GO:0044267)	cellular protein metabolic process
(GO:0045333)	cellular respiration
(GO:0007098)	centrosome cycle
(GO:0051298)	centrosome duplication
(GO:0051297)	centrosome organization
(GO:0002181)	cytoplasmic translation
(GO:0022900)	electron transport chain
(GO:0015980)	energy derivation by oxidation of organic compounds
(GO:0010467)	gene expression
(GO:0006091)	generation of precursor metabolites and energy
(GO:0009059)	macromolecule biosynthetic process
(GO:0043170)	macromolecule metabolic process
(GO:0008152)	metabolic process
(GO:0000226)	microtubule cytoskeleton organization
(GO:0031023)	microtubule organizing center organization
(GO:0042775)	mitochondrial ATP synthesis coupled electron transport
(GO:0006120)	mitochondrial electron transport, NADH to ubiquinone
(GO:0032543)	mitochondrial translation
(GO:0006807)	nitrogen compound metabolic process
(GO:0009116)	nucleoside metabolic process
(GO:0009123)	nucleoside monophosphate metabolic process
(GO:0006753)	nucleoside phosphate metabolic process
(GO:0009141)	nucleoside triphosphate metabolic process
(GO:0009117)	nucleotide metabolic process
(GO:0006996)	organelle organization
(GO:1901576)	organic substance biosynthetic process
(GO:0071704)	organic substance metabolic process
(GO:1901566)	organonitrogen compound biosynthetic process
(GO:1901564)	organonitrogen compound metabolic process
(GO:0006119)	oxidative phosphorylation
(GO:0043043)	peptide biosynthetic process
(GO:0006518)	peptide metabolic process
(GO:0044238)	primary metabolic process
(GO:0019538)	protein metabolic process
(GO:0042278)	purine nucleoside metabolic process
(GO:0009126)	purine nucleoside monophosphate metabolic process
(GO:0009144)	purine nucleoside triphosphate metabolic process
(GO:0006163)	purine nucleotide metabolic process
(GO:0046128)	purine ribonucleoside metabolic process
(GO:0009167)	purine ribonucleoside monophosphate metabolic process
(GO:0009205)	purine ribonucleoside triphosphate metabolic process
(GO:0009150)	purine ribonucleotide metabolic process
(GO:0022904)	respiratory electron transport chain
(GO:0009119)	ribonucleoside metabolic process
(GO:0009161)	ribonucleoside monophosphate metabolic process
(GO:0009199)	ribonucleoside triphosphate metabolic process

(GO:0009259)	ribonucleotide metabolic process
(GO:0019693)	ribose phosphate metabolic process
(GO:0006412)	translation

Supplementary Table 3.2 Gene Ontology terms enriched in yellow

(GO:0030036)	actin cytoskeleton organization
(GO:0030029)	actin filament-based process
(GO:0034333)	adherens junction assembly
(GO:0034332)	adherens junction organization
(GO:0001667)	ameboid-type cell migration
(GO:0048856)	anatomical structure development
(GO:0048646)	anatomical structure formation involved in morphogenesis
(GO:0009653)	anatomical structure morphogenesis
(GO:0048513)	animal organ development
(GO:0009887)	animal organ morphogenesis
(GO:0009952)	anterior/posterior pattern specification
(GO:0043297)	apical junction assembly
(GO:0048736)	appendage development
(GO:0035107)	appendage morphogenesis
(GO:0008356)	asymmetric cell division
(GO:0055059)	asymmetric neuroblast division
(GO:0008105)	asymmetric protein localization
(GO:0098722)	asymmetric stem cell division
(GO:0009798)	axis specification
(GO:0061564)	axon development
(GO:0007411)	axon guidance
(GO:0007409)	axonogenesis
(GO:0007610)	behavior
(GO:0022610)	biological adhesion
(GO:0065007)	biological regulation
(GO:0008150)	biological process
(GO:0007350)	blastoderm segmentation
(GO:0007298)	border follicle cell migration
(GO:0007155)	cell adhesion
(GO:0061343)	cell adhesion involved in heart morphogenesis
(GO:0007154)	cell communication
(GO:0048468)	cell development
(GO:0030154)	cell differentiation
(GO:0051301)	cell division
(GO:0045165)	cell fate commitment
(GO:0016049)	cell growth
(GO:0034329)	cell junction assembly
(GO:0034330)	cell junction organization
(GO:0048469)	cell maturation
(GO:0016477)	cell migration
(GO:0000902)	cell morphogenesis
(GO:0000904)	cell morphogenesis involved in differentiation
(GO:0048667)	cell morphogenesis involved in neuron differentiation
(GO:0048870)	cell motility
(GO:0032990)	cell part morphogenesis
(GO:0048858)	cell projection morphogenesis
(GO:0030030)	cell projection organization
(GO:0008283)	cell proliferation
(GO:0008037)	cell recognition
(GO:0007166)	cell surface receptor signaling pathway
(GO:0098609)	cell-cell adhesion
(GO:0098742)	cell-cell adhesion via plasma-membrane adhesion molecules
(GO:0007043)	cell-cell junction assembly
(GO:0045216)	cell-cell junction organization
(GO:0007267)	cell-cell signaling
(GO:0022607)	cellular component assembly

(GO:0044085)	cellular component biogenesis
(GO:0032989)	cellular component morphogenesis
(GO:0016043)	cellular component organization
(GO:0071840)	cellular component organization or biogenesis
(GO:0048869)	cellular developmental process
(GO:0051641)	cellular localization
(GO:0044265)	cellular macromolecule catabolic process
(GO:0044260)	cellular macromolecule metabolic process
(GO:0044237)	cellular metabolic process
(GO:0009987)	cellular process
(GO:0022412)	cellular process involved in reproduction in multicellular organism
(GO:0044267)	cellular protein metabolic process
(GO:0006464)	cellular protein modification process
(GO:0051716)	cellular response to stimulus
(GO:0007417)	central nervous system development
(GO:0006935)	chemotaxis
(GO:0072359)	circulatory system development
(GO:0050890)	cognition
(GO:0002066)	columnar/cuboidal epithelial cell development
(GO:0002065)	columnar/cuboidal epithelial cell differentiation
(GO:0048749)	compound eye development
(GO:0001745)	compound eye morphogenesis
(GO:0046667)	compound eye retinal cell programmed cell death
(GO:0007010)	cytoskeleton organization
(GO:0048589)	developmental growth
(GO:0021700)	developmental maturation
(GO:0032502)	developmental process
(GO:0003006)	developmental process involved in reproduction
(GO:0055123)	digestive system development
(GO:0048565)	digestive tract development
(GO:0007391)	dorsal closure
(GO:0009790)	embryo development
(GO:0009792)	embryo development ending in birth or egg hatching
(GO:0000578)	embryonic axis specification
(GO:0001700)	embryonic development via the syncytial blastoderm
(GO:0048598)	embryonic morphogenesis
(GO:0009880)	embryonic pattern specification
(GO:0006897)	endocytosis
(GO:0007167)	enzyme linked receptor protein signaling pathway
(GO:0002064)	epithelial cell development
(GO:0030855)	epithelial cell differentiation
(GO:0010631)	epithelial cell migration
(GO:0060562)	epithelial tube morphogenesis
(GO:0060429)	epithelium development
(GO:0090132)	epithelium migration
(GO:0030010)	establishment of cell polarity
(GO:0051234)	establishment of localization
(GO:0051649)	establishment of localization in cell
(GO:0051656)	establishment of organelle localization
(GO:0001736)	establishment of planar polarity
(GO:0007164)	establishment of tissue polarity
(GO:0007163)	establishment or maintenance of cell polarity
(GO:0001654)	eye development
(GO:0048592)	eye morphogenesis
(GO:0001754)	eye photoreceptor cell differentiation
(GO:0007292)	female gamete generation
(GO:0007276)	gamete generation
(GO:0048699)	generation of neurons
(GO:0007281)	germ cell development
(GO:0007390)	germ-band shortening
(GO:0030718)	germ-line stem cell population maintenance
(GO:0042063)	gliogenesis
complete	GO biological process

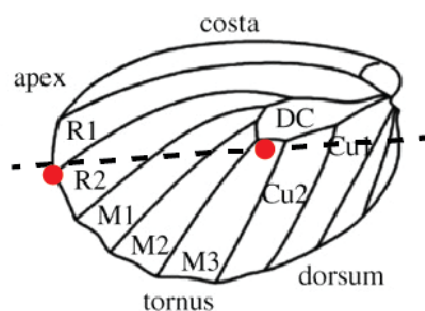
(GO:0040007)	growth
(GO:0007507)	heart development
(GO:0003007)	heart morphogenesis
(GO:0007444)	imaginal disc development
(GO:0007560)	imaginal disc morphogenesis
(GO:0048737)	imaginal disc-derived appendage development
(GO:0035114)	imaginal disc-derived appendage morphogenesis
(GO:0007476)	imaginal disc-derived wing morphogenesis
(GO:0008586)	imaginal disc-derived wing vein morphogenesis
(GO:0002168)	instar larval development
(GO:0002165)	instar larval or pupal development
(GO:0048707)	instar larval or pupal morphogenesis
(GO:0035556)	intracellular signal transduction
(GO:0046907)	intracellular transport
(GO:0002164)	larval development
(GO:0007611)	learning or memory
(GO:0051179)	localization
(GO:0051674)	localization of cell
(GO:0040011)	locomotion
(GO:0033036)	macromolecule localization
(GO:0043412)	macromolecule modification
(GO:0098727)	maintenance of cell number
(GO:0061024)	membrane organization
(GO:0007613)	memory
(GO:0007552)	metamorphosis
(GO:0001738)	morphogenesis of a polarized epithelium
(GO:0002009)	morphogenesis of an epithelium
(GO:0016331)	morphogenesis of embryonic epithelium
(GO:0008045)	motor neuron axon guidance
(GO:0006928)	movement of cell or subcellular component
(GO:0051704)	multi-organism process
(GO:0044703)	multi-organism reproductive process
(GO:0007275)	multicellular organism development
(GO:0032504)	multicellular organism reproduction
(GO:0032501)	multicellular organismal process
(GO:0048609)	multicellular organismal reproductive process
(GO:0042692)	muscle cell differentiation
(GO:0007517)	muscle organ development
(GO:0061061)	muscle structure development
(GO:0048519)	negative regulation of biological process
(GO:0010648)	negative regulation of cell communication
(GO:0010721)	negative regulation of cell development
(GO:0045596)	negative regulation of cell differentiation
(GO:0051129)	negative regulation of cellular component organization
(GO:0031324)	negative regulation of cellular metabolic process
(GO:0048523)	negative regulation of cellular process
(GO:0032269)	negative regulation of cellular protein metabolic process
(GO:0051093)	negative regulation of developmental process
(GO:1902532)	negative regulation of intracellular signal transduction
(GO:0033673)	negative regulation of kinase activity
(GO:0010605)	negative regulation of macromolecule metabolic process
(GO:0009892)	negative regulation of metabolic process
(GO:0044092)	negative regulation of molecular function
(GO:0051241)	negative regulation of multicellular organismal process
(GO:0051961)	negative regulation of nervous system development
(GO:0045936)	negative regulation of phosphate metabolic process
(GO:0010563)	negative regulation of phosphorus metabolic process
(GO:0042326)	negative regulation of phosphorylation
(GO:0006469)	negative regulation of protein kinase activity
(GO:0051248)	negative regulation of protein metabolic process
(GO:0031400)	negative regulation of protein modification process
(GO:0001933)	negative regulation of protein phosphorylation
(GO:0048585)	negative regulation of response to stimulus

(GO:0009968)	negative regulation of signal transduction
(GO:0023057)	negative regulation of signaling
(GO:0051348)	negative regulation of transferase activity
(GO:0007399)	nervous system development
(GO:0061351)	neural precursor cell proliferation
(GO:0055057)	neuroblast division
(GO:0007405)	neuroblast proliferation
(GO:0022008)	neurogenesis
(GO:0050877)	neurological system process
(GO:0048666)	neuron development
(GO:0030182)	neuron differentiation
(GO:0031175)	neuron projection development
(GO:0097485)	neuron projection guidance
(GO:0048812)	neuron projection morphogenesis
(GO:0008038)	neuron recognition
(GO:0036445)	neuronal stem cell division
(GO:0048477)	oogenesis
(GO:0007424)	open tracheal system development
(GO:0051640)	organelle localization
(GO:0006996)	organelle organization
(GO:0030707)	ovarian follicle cell development
(GO:0007297)	ovarian follicle cell migration
(GO:0007389)	pattern specification process
(GO:0007422)	peripheral nervous system development
(GO:0006909)	phagocytosis
(GO:0006911)	phagocytosis, engulfment
(GO:0006796)	phosphate-containing compound metabolic process
(GO:0006793)	phosphorus metabolic process
(GO:0016310)	phosphorylation
(GO:0046530)	photoreceptor cell differentiation
(GO:0048518)	positive regulation of biological process
(GO:0010647)	positive regulation of cell communication
(GO:0051130)	positive regulation of cellular component organization
(GO:0031325)	positive regulation of cellular metabolic process
(GO:0048522)	positive regulation of cellular process
(GO:0032270)	positive regulation of cellular protein metabolic process
(GO:0051094)	positive regulation of developmental process
(GO:0010604)	positive regulation of macromolecule metabolic process
(GO:0009893)	positive regulation of metabolic process
(GO:0051240)	positive regulation of multicellular organismal process
(GO:0051962)	positive regulation of nervous system development
(GO:0051247)	positive regulation of protein metabolic process
(GO:0031401)	positive regulation of protein modification process
(GO:0048584)	positive regulation of response to stimulus
(GO:0023056)	positive regulation of signaling
(GO:0009886)	post-embryonic animal morphogenesis
(GO:0048569)	post-embryonic animal organ development
(GO:0048563)	post-embryonic animal organ morphogenesis
(GO:0035120)	post-embryonic appendage morphogenesis
(GO:0009791)	post-embryonic development
(GO:0008104)	protein localization
(GO:0019538)	protein metabolic process
(GO:0036211)	protein modification process
(GO:0006468)	protein phosphorylation
(GO:0045466)	R7 cell differentiation
(GO:0003002)	regionalization
(GO:0032956)	regulation of actin cytoskeleton organization
(GO:0032970)	regulation of actin filament-based process
(GO:0022603)	regulation of anatomical structure morphogenesis
(GO:0090066)	regulation of anatomical structure size
(GO:0050789)	regulation of biological process
(GO:0065008)	regulation of biological quality
(GO:0009889)	regulation of biosynthetic process

(GO:0050790)	regulation of catalytic activity
(GO:0010646)	regulation of cell communication
(GO:0060284)	regulation of cell development
(GO:0045595)	regulation of cell differentiation
(GO:0031344)	regulation of cell projection organization
(GO:0034248)	regulation of cellular amide metabolic process
(GO:0031326)	regulation of cellular biosynthetic process
(GO:0044087)	regulation of cellular component biogenesis
(GO:0051128)	regulation of cellular component organization
(GO:0032535)	regulation of cellular component size
(GO:0060341)	regulation of cellular localization
(GO:2000112)	regulation of cellular macromolecule biosynthetic process
(GO:0031323)	regulation of cellular metabolic process
(GO:0050794)	regulation of cellular process
(GO:0032268)	regulation of cellular protein metabolic process
(GO:0080135)	regulation of cellular response to stress
(GO:0051493)	regulation of cytoskeleton organization
(GO:0048638)	regulation of developmental growth
(GO:0050793)	regulation of developmental process
(GO:0010468)	regulation of gene expression
(GO:0040008)	regulation of growth
(GO:0045610)	regulation of hemocyte differentiation
(GO:1902531)	regulation of intracellular signal transduction
(GO:0043549)	regulation of kinase activity
(GO:0032879)	regulation of localization
(GO:0010556)	regulation of macromolecule biosynthetic process
(GO:0060255)	regulation of macromolecule metabolic process
(GO:0043408)	regulation of MAPK cascade
(GO:0019222)	regulation of metabolic process
(GO:0065009)	regulation of molecular function
(GO:2000026)	regulation of multicellular organismal development
(GO:0051239)	regulation of multicellular organismal process
(GO:0051960)	regulation of nervous system development
(GO:0050767)	regulation of neurogenesis
(GO:1904396)	regulation of neuromuscular junction development
(GO:0045664)	regulation of neuron differentiation
(GO:0051171)	regulation of nitrogen compound metabolic process
(GO:1903506)	regulation of nucleic acid-templated transcription
(GO:0019219)	regulation of nucleobase-containing compound metabolic process
(GO:0033043)	regulation of organelle organization
(GO:0019220)	regulation of phosphate metabolic process
(GO:0051174)	regulation of phosphorus metabolic process
(GO:0042325)	regulation of phosphorylation
(GO:0080090)	regulation of primary metabolic process
(GO:0045859)	regulation of protein kinase activity
(GO:0051246)	regulation of protein metabolic process
(GO:0031399)	regulation of protein modification process
(GO:0001932)	regulation of protein phosphorylation
(GO:0048583)	regulation of response to stimulus
(GO:0080134)	regulation of response to stress
(GO:2001141)	regulation of RNA biosynthetic process
(GO:0051252)	regulation of RNA metabolic process
(GO:0009966)	regulation of signal transduction
(GO:0023051)	regulation of signaling
(GO:0051963)	regulation of synapse assembly
(GO:0050807)	regulation of synapse organization
(GO:0050803)	regulation of synapse structure or activity
(GO:0008582)	regulation of synaptic growth at neuromuscular junction
(GO:0006357)	regulation of transcription from RNA polymerase II promoter
(GO:0006355)	regulation of transcription, DNA-templated
(GO:0051338)	regulation of transferase activity
(GO:0051049)	regulation of transport
(GO:0035152)	regulation of tube architecture, open tracheal system

(GO:0072001)	renal system development
(GO:0000003)	reproduction
(GO:0022414)	reproductive process
(GO:0048608)	reproductive structure development
(GO:0061458)	reproductive system development
(GO:0060541)	respiratory system development
(GO:0042221)	response to chemical
(GO:0009605)	response to external stimulus
(GO:0050896)	response to stimulus
(GO:0046666)	retinal cell programmed cell death
(GO:0035282)	segmentation
(GO:0007423)	sensory organ development
(GO:0090596)	sensory organ morphogenesis
(GO:0019991)	septate junction assembly
(GO:0019953)	sexual reproduction
(GO:0007165)	signal transduction
(GO:0023052)	signaling
(GO:0098602)	single organism cell adhesion
(GO:0044702)	single organism reproductive process
(GO:0044700)	single organism signaling
(GO:0016337)	single organismal cell-cell adhesion
(GO:0044707)	single-multicellular organism process
(GO:0044708)	single-organism behavior
(GO:1902580)	single-organism cellular localization
(GO:0044763)	single-organism cellular process
(GO:0044767)	single-organism developmental process
(GO:1902578)	single-organism localization
(GO:1902589)	single-organism organelle organization
(GO:0044699)	single-organism process
(GO:0044765)	single-organism transport
(GO:0007525)	somatic muscle development
(GO:0048103)	somatic stem cell division
(GO:0048863)	stem cell differentiation
(GO:0017145)	stem cell division
(GO:0019827)	stem cell population maintenance
(GO:0072089)	stem cell proliferation
(GO:0097435)	supramolecular fiber organization
(GO:0048731)	system development
(GO:0003008)	system process
(GO:0042330)	taxis
(GO:0009888)	tissue development
(GO:0090130)	tissue migration
(GO:0048729)	tissue morphogenesis
(GO:0007169)	transmembrane receptor protein tyrosine kinase signaling pathway
(GO:0006810)	transport
(GO:0035295)	tube development
(GO:0035239)	tube morphogenesis
(GO:0001655)	urogenital system development
(GO:0016192)	vesicle-mediated transport
(GO:0035220)	wing disc development
(GO:0007472)	wing disc morphogenesis
(GO:0035222)	wing disc pattern formation

Figure S3.1: Hindwing dissection diagram. Pupal hindwings were cut into approximate Anterior and Posterior pieces; the cut was chosen to keep the region of the wing that forms the yellow band entirely on one side of the cut. While no pattern is visible yet in pupal wings, the venation pattern can clearly be seen. The key landmarks used are highlighted with red circles; these were the posterior base of the discal cell (DC), and the point where the R1 vein contacts the wing edge.



References

- ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31, 166-9.
- BELDADE, P., SAENKO, S. V., PUL, N. & LONG, A. D. 2009. A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome. *PLoS Genet*, 5, e1000366.
- BLUM, M. J. 2002. Rapid movement of a *Heliconius* hybrid zone: evidence for phase III of Wright's shifting balance theory? *Evolution*, 56, 1992-8.
- BROWN, S., HU, N. & HOMBRIA, J. C. 2001. Identification of the first invertebrate interleukin JAK/STAT receptor, the *Drosophila* gene domeless. *Curr Biol*, 11, 1700-5.
- CHALLIS, R. J., KUMAR, S., DASMAHAPATRA, K. K. K., JIGGINS, C. D. & BLAXTER, M. 2016. Lepbase: the Lepidopteran genome database. *BioRxiv*, 056994.
- DERIVERY, E. & GAUTREAU, A. 2010. Evolutionary conservation of the WASH complex, an actin polymerization machine involved in endosomal fission. *Commun Integr Biol*, 3, 227-30.
- FATICA, A. & BOZZONI, I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews. Genetics*, 15, 7.
- FERGUSON, L., LEE, S. F., CHAMBERLAIN, N., NADEAU, N., JORON, M., BAXTER, S., WILKINSON, P., PAPANICOLAOU, A., KUMAR, S., KEE, T. J., CLARK, R., DAVIDSON, C., GLITHERO, R., BEASLEY, H., VOGEL, H., FFRENCH-CONSTANT, R. & JIGGINS, C. 2010. Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol Ecol*, 19 Suppl 1, 240-54.
- ITO, K., KATSUMA, S., KUWAZAKI, S., JOURAKU, A., FUJIMOTO, T., SAHARA, K., YASUKOCHI, Y., YAMAMOTO, K., TABUNOKI, H., YOKOYAMA, T., KADONO-OKUDA, K. & SHIMADA, T. 2016. Mapping and recombination analysis of two moth colour mutations, Black moth and Wild wing spot, in the silkworm *Bombyx mori*. *Heredity (Edinb)*, 116, 52-9.
- JORON, M., PAPA, R., BELTRAN, M., CHAMBERLAIN, N., MAVAREZ, J., BAXTER, S., ABANTO, M., BERMINGHAM, E., HUMPHRAY, S. J., ROGERS, J., BEASLEY, H., BARLOW, K., FFRENCH-CONSTANT, R. H., MALLET, J., MCMILLAN, W. O. & JIGGINS, C. D. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol*, 4, e303.
- KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 12, 357-60.
- LAVOIE, C. A., PLATT, R. N., 2ND, NOVICK, P. A., COUNTERMAN, B. A. & RAY, D. A. 2013. Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera. *Mob DNA*, 4, 21.
- LI, X., FAN, D., ZHANG, W., LIU, G., ZHANG, L., ZHAO, L., FANG, X., CHEN, L., DONG, Y., CHEN, Y., DING, Y., ZHAO, R., FENG, M., ZHU, Y., FENG, Y., JIANG, X., ZHU, D., XIANG, H., FENG, X., LI, S., WANG, J., ZHANG, G., KRONFORST, M. R. & WANG, W. 2015. Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat Commun*, 6, 8212.
- MALLET, J. 1986. Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning colour clines. *Heredity*.
- MALLET, J. 1993. Speciation, raiation, and color pattern evolution in *Heliconius* butterflies: evidence from hybrid zones. *Hybrid zones and the evolutionary process*, 226-260.

- MAROJA, L. S., ALSCHULER, R., MCMILLAN, W. O. & JIGGINS, C. D. 2012. Partial complementarity of the mimetic yellow bar phenotype in *Heliconius* butterflies. *PLoS One*, 7, e48627.
- MARTIN, A. & ORGOGOZO, V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*, 67, 1235-50.
- MARTIN, A., PAPA, R., NADEAU, N. J., HILL, R. I., COUNTERMAN, B. A., HALDER, G., JIGGINS, C. D., KRONFORST, M. R., LONG, A. D., MCMILLAN, W. O. & REED, R. D. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proc Natl Acad Sci U S A*, 109, 12632-7.
- MARTIN, A. & REED, R. D. 2014. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Dev Biol*, 395, 367-78.
- MARTIN, A., SERANO, J. M., JARVIS, E., BRUCE, H. S., WANG, J., RAY, S., BARKER, C. A., O'CONNELL, L. C. & PATEL, N. H. 2016. CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Curr Biol*, 26, 14-26.
- MCLYSAGHT, A. & GUERZONI, D. 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philos Trans R Soc Lond B Biol Sci*, 370, 20140332.
- NADEAU, N. J., MARTIN, S. H., KOZAK, K. M., SALAZAR, C., DASMAHAPATRA, K. K., DAVEY, J. W., BAXTER, S. W., BLAXTER, M. L., MALLET, J. & JIGGINS, C. D. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Mol Ecol*, 22, 814-26.
- NADEAU, N. J., PARDO-DIAZ, C., WHIBLEY, A., SUPPLE, M. A., SAENKO, S. V., WALLBANK, R. W., WU, G. C., MAROJA, L., FERGUSON, L., HANLY, J. J., HINES, H., SALAZAR, C., MERRILL, R. M., DOWLING, A. J., FFRENCH-CONSTANT, R. H., LLAURENS, V., JORON, M., MCMILLAN, W. O. & JIGGINS, C. D. 2016. The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, 534, 106-10.
- NADEAU, N. J., RUIZ, M., SALAZAR, P., COUNTERMAN, B., MEDINA, J. A., ORTIZ-ZUAZAGA, H., MORRISON, A., MCMILLAN, W. O., JIGGINS, C. D. & PAPA, R. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res*, 24, 1316-33.
- NADEAU, N. J., WHIBLEY, A., JONES, R. T., DAVEY, J. W., DASMAHAPATRA, K. K., BAXTER, S. W., QUAIL, M. A., JORON, M., FFRENCH-CONSTANT, R. H., BLAXTER, M. L., MALLET, J. & JIGGINS, C. D. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci*, 367, 343-53.
- PACE, R. M., GRBIC, M. & NAGY, L. M. 2016. Composition and genomic organization of arthropod Hox clusters. *Evodevo*, 7, 11.
- PALMEIRIM, I., HENRIQUE, D., ISH-HOROWICZ, D. & POURQUIÉ, O. 1997. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*, 91, 639-648.
- RODRIGUEZ-MESA, E., ABREU-BLANCO, M. T., ROSALES-NIEVES, A. E. & PARKHURST, S. M. 2012. Developmental expression of *Drosophila* Wiskott-Aldrich Syndrome family proteins. *Dev Dyn*, 241, 608-26.
- SAENKO, S. V., BRAKEFIELD, P. M. & BELDADE, P. 2010. Single locus affects embryonic segment polarity and multiple aspects of an adult evolutionary novelty. *BMC Biol*, 8, 111.
- SHEPPARD, P., TURNER, J., BROWN, K., BENSON, W. & SINGER, M. 1985. Genetics and the evolution of Muellierian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 433-610.

SURRIDGE, A. K., LOPEZ-GOMOLLON, S., MOXON, S., MAROJA, L. S., RATHJEN, T., NADEAU, N. J., DALMAY, T. & JIGGINS, C. D. 2011. Characterisation and expression of microRNAs in developing wings of the neotropical butterfly *Heliconius melpomene*. *BMC Genomics*, 12, 62.

TOLIC-NORRELYKKE, S. F., ENGH, A. M., LANDICK, R. & GELLES, J. 2004. Diversity in the rates of transcript elongation by single RNA polymerase molecules. *J Biol Chem*, 279, 3292-9.

TOMANCAK, P., BEATON, A., WEISZMANN, R., KWAN, E., SHU, S., LEWIS, S. E., RICHARDS, S., ASHBURNER, M., HARTENSTEIN, V., CELNIKER, S. E. & RUBIN, G. M. 2002. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol*, 3, RESEARCH0088.

VAN BELLEGHEM, S. M., RASTAS, P., PAPANICOLAOU, A., MARTIN, S. H., ARIAS, C. F., SUPPLE, M. A., HANLY, J. J., MALLET, J., LEWIS, J. J., HINES, H. M., RUIZ, M., SALAZAR, C., LINARES, M., MOREIRA, G. R. P., JIGGINS, C. D., COUNTERMAN, B. A., MCMILLAN, W. O. & PAPA, R. 2017. Complex modular architecture around a simple toolkit of wing pattern genes. *Nat Ecol Evol*, 1.

VAN'T HOF, A. E., CAMPAGNE, P., RIGDEN, D. J., YUNG, C. J., LINGLEY, J., QUAIL, M. A., HALL, N., DARBY, A. C. & SACCHERI, I. J. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, 534, 102-5.

Chapter 4

The gene regulatory landscape of the butterfly wing – a molecular dissection of the Nymphalid Ground Plan

ABSTRACT

A major challenge to evolutionary developmental biology is to understand the how modifications to gene regulatory networks can lead to biological diversity. *Heliconius* butterfly wing patterns provide an excellent example of this diversity. I described the expression of genes in the developing wings of three species of butterfly, detecting differential expression of key developmental factors involved in determining the identity, axes and morphology of insect wings. Many of these factors, along with others not previously described as being expressed in developing wings, were expressed in conserved domains of expression in all three species. Other factors, most notably the constituents of the Wnt signalling pathway, varied in different lineages with wing pattern. A deeper understanding of factors that are expressed in the wing in correlation with pattern elements will assist in decoding the regulatory linkages that lead to the differential expression of the switch genes *optix*, *WntA* and *cortex*, and will hopefully be of use to the general understanding of butterfly wing pattern evolution.

I: INTRODUCTION

Evolution of *cis*-regulatory elements has been proposed as a primary mechanism of evolution of form because *cis*-modifications can create discrete tissue specific changes without deleterious pleiotropic effects (Stern and Orgogozo, 2009, Carroll, 2000, Gompel and Prud'homme, 2009). Modifications of spatial patterns of gene expression can occur by such *cis*-regulatory modifications. For example, the gene *En* is a deeply conserved component of arthropod segmentation which specifies the posterior compartment (Patel et al., 1989) and is expressed in this pattern in the developing wings of *Drosophila* species. Some *melanogaster*-group males have a melanic spot on the anterior tip of the wing. The gene *yellow*, which is associated with

melanin synthesis, has gained regulatory linkage with *en*, which acts as a repressive cue and sculpts the posterior boundary of the melanic spot. Here, *engrailed* has gained a new function without any change to its activity, protein sequence or expression, simply because it was expressed in a spatiotemporal pattern that could be innovated upon and co-opted for a novel function.

In contrast, it is also possible for such expression differences to arise due to changes in spatial arrangement of *trans*-acting regulatory factors. For example, the species *Drosophila guttifera* has a complex wing pattern including melanic patches associated with crossveins and campaniform sensilla. The gene *Wingless*, usually expressed at the wing margin in developing wings, is capable of activating *yellow* to induce melanin deposition, and has been redeployed to be expressed in the campaniform sensilla via *cis*-regulatory modifications (Werner et al., 2010).

It has been theorised that transcription factors are likely targets of *cis*-regulatory modifications during the evolution of morphology and form. However, in *Heliconius* for example, the main patterning loci are *optix*, a transcription factor, *WntA*, a signalling ligand, and one or more of *cortex*, a cell cycle regulator, *domeless*, a receptor tyrosine kinase or *washout*, an actin binding protein; while transcription factors are capable of playing a special role, a picture is emerging that other types of genes may be capable of fulfilling this role, in particular ligands (Martin and Orgogozo, 2017).

In earlier chapters I have described the genetic and functional basis of the yellow bar with transcriptomic methods and identified regulatory modules associated with the red forewing band through association and phylogenetic topology. I hypothesize that the set of developmental decisions taken by differentiating scale precursors are spatially patterned by factors that are already present in the wing, and which are required for normal development. However, our picture of what factors are differentially expressed in developing butterfly wings, and the timing of their expression relative to the differentiation of scale precursor cells, is limited and largely guided by inference from model systems, especially *Drosophila*. A fuller understanding of how these factors relate to butterfly wing pattern requires a better understanding of spatial and temporal expression in developing butterfly wings.

Here, I will discuss the historical frameworks for understanding butterfly wing pattern formation and evolution and relate these to our present understanding of wing development. In order to improve our butterfly-specific knowledge of wing development, I have carried out multispecies comparative transcriptomic sequencing in order to investigate which factors are expressed in association with wing pattern, but which will also improve general understanding of wing pattern development in Lepidoptera.

The Nymphalid Ground Plan

In seeking to understand the evolution of animal body forms, scientific workers from the late 19th and early 20th century began to focus on obvious homologies between structures in different lineages, latterly incorporating data from fossils. Perhaps the most notable example is the examination of the vertebrate limb, and the eventual understanding that sarcopterygian appendages – both the fins of lobe-finned fishes and the limbs of tetrapods – are homologous structures built from similar homologous components, and with a shared common ancestor (Braus, 1900, Pierce et al., 2013). It was in this context that during the 1920s, two German entomologists independently developed the idea that the wing patterns of Lepidoptera – especially the Nymphalidae – contained many examples of elements that appear homologous. Both Schwanwitsch (1924) and Suffert (1927) constructed remarkably similar models, involving sets of concentric pattern elements moving outwards from the proximal wing hinge to the distal margin, referred to as “symmetry systems”. These models have become known as the Nymphalid Ground Plan (NGP), and have been elaborated at length by Fred Nijhout, most fully summarised in his 1991 book *The Development and Evolution of Butterfly Wing Patterns* (Nijhout, 1991).

Nijhout argues that, unlike the homology seen in appendages, the NGP does not represent common descent from some ancestral state, but is rather a list of homologies which represents “the maximum pattern that could be present in a generalised butterfly”. He also argues that most (if not all) patterns can be explained by diffusion of factors from wing veins, illustrated by frequent dislocation of elements within wing cells in many species. Through extensive morphometric measurements and modelling of repeated intervein elements, he argues that evolution between different types

within genera can be explained by a small number of mutational steps in parameters including signal diffusion rate and stability, and by reaction-diffusion interactions.

If the NGP does not represent common descent from an ancestral state, then what does it represent? Developmental constraint and bias are likely to influence the evolution of wing pattern. Any particular pattern phenotype can be conceptualised as a point on an adaptive landscape. There are peaks of fitness within this landscape, which may correspond to more effective crypsis, more effective mimicry, or more effective signalling to mates. A mutation that causes a change to wing pattern will cause movement through the landscape. There are two primary forces that shape this adaptive landscape. The first is selection. A mutational change that leads to better mimicry, say, corresponds to a step up a slope on the adaptive landscape, and ultimately such changes will be selected for and spread to fixation in a population. A mutational change that reduces camouflage in a cryptic species corresponds to a descent of an adaptive peak and reduced fitness. In this way, selection can steer wing patterns to converge, even in diverse lineages. Concentric rings of dark and light scales look like eyes to potential predators, and this induces a startle response which reduces predation rate. This selective pressure may lead to the convergent evolution of concentric rings in many lineages with no need to invoke common descent. The second force is developmental bias and constraint. Some parts of the adaptive landscape may be easier to reach than others; for example, there may be n different mutations that can produce pattern A, but only $1/n$ mutations that can produce pattern B, meaning pattern A is more likely to recur. Likewise, some imaginable patterns are not producible, i.e. there are no mutational changes that can generate that given pattern in that given system, and so that pattern cannot occur. Such biases and constraints may come from physical and physiological properties of wings (a 2D surface cannot produce a 3D pattern, say) but in the case of butterfly wings could also come from the gene regulatory network that forms the wing.

Nijhout dubs these developmental factors ‘developmental space’, and argues that the specific parameters of developmental space may vary by degree between species, and that this will account for some pattern evolution. However, largely we might expect developmental space in the wing to have some highly conserved aspects. Nijhout argues that the interplay of selection and developmental space ultimately leads to the

observed homology between butterfly wing pattern elements, although he ultimately puts more emphasis on the role of developmental space than on selection.

Heliconius wing patterns, with their bold patches that cut across wing veins, do not obviously follow the NGP. Nijhout acknowledges that for the NGP model to adequately explain these *Heliconius* wing patterns, it must be “somewhat different” from the generalised NGP – though some patterns, such as marginal bands and hindwing rays, confound the model less than others (Nijhout et al., 1990, Nijhout and Wray, 1988). It should also be noted that despite Nijhout’s emphasis on the role of veins as sources and sinks of signal in the NGP model, vein-less mutants have been observed on multiple occasions in insectary stocks of different *Heliconius* species, notably by Reed et al (Reed and Gilbert, 2004) and Southcott (unpublished) in which wing pattern is broadly unaffected, and a mutant *Papilio xuthus* with reduced veins has been described, which loses some pattern detail but retains the broad pattern domains (Koch and Nijhout, 2002). Alternative attempts to develop a model for the development of wing pattern in *Heliconius* were proposed by Larry Gilbert.

Gilbert’s Shutters and Windows

The conceptual starting point for Gilbert’s interpretation of wing pattern evolution is very different to that of the NGP. He began with the observation that the high phenotypic diversity, both between and within species of *Heliconius*, implies that there must be some special distinguishing feature of this clade. While the material for investigation of the NGP was the naturally occurring species-level variation found in nature, Gilbert encouraged interspecies and inter-race crosses to occur in his extensive insectary stocks, and from the diversity of form that arose he was able to build a morphospace of the ranges of patterns that can possibly be achieved within *Heliconius* (though mainly focusing on the *melpomene-cydno-silvaniform* clade, which is where most extant and ancient introgression is observed (Gilbert, 2003). The approach used by Gilbert of actively recombining wing pattern genes into different backgrounds could potentially separate bias caused by selection from inherent developmental constraint, as the range of aberrant patterns do not face the challenge of selection before they enter experimental observations.

Gilbert observed that, when crossed into different species and subspecies backgrounds, some wing pattern genes consistently modify the pattern in the same way, irrespective of the genomic background. He terms these genes ‘toolbox genes’. Other genes had broader effects on pattern, causing the presence or absence of one colour, but these effects did vary in differing genomic contexts. He terms these genes ‘switch genes’. He conceptualises the interactions as a window with shutters.

For example, an allele linked to a yellow rectangle on the hindwing of *H. pachinus* was crossed onto a *H. cydno* background, where it caused a large yellow region. This gene, he argues, is acting as a general ‘on’ switch. In addition to this general on-switch, *H. pachinus* has an additional toolbox-gene that shapes this pattern. In other words, the switch gene turns the ‘window’ yellow, and then the toolbox gene acts as a ‘shutter’, covering portions of this window. These windows, which are competent to respond to switch signals, are bounded by ‘walls’ – for example, the distal tip of the forewing is consistently black, and no switch or toolbox gene will extend into this region.

Gilbert does not attempt to integrate his model with the genetics and linkage analysis performed by Sheppard and others (Sheppard et al., 1985). He also disputed the idea that there are sets of tightly linked loci (i.e. Table 1.1), arguing that the observation of linkage can be explained by multiple unlinked interacting loci in epistasis. This is demonstrably not true.

One particularly notable feature of this model is that it is able to account for the generation of novel diversity by introgression. In particular, he describes an insectary cross of Costa Rican *H. ismenius* with Ecuadorian *H. cydno*, which produced F1 offspring with wing patterns that could credibly mimic Ecuadorian *H. hecalesia*. While such an event would be extremely rare, it may permit a population to enter a new Müllerian mimicry ring much more rapidly than via gradual substitution of new mutant alleles. A similar model has since been proposed for the evolution of the wing pattern of *H. heurippa* from introgression between *H. t. linarezi* and *H. m. melpomene*, indicating the productiveness of this aspect of the hypothesis (Brown and Mielke, 1972, Salazar et al., 2010).

An alternative view?

Both the NGP model and the Windows and Shutters model describe wing pattern – though in different ways - and also draw inferences and predictions about the evolution and development of butterfly wing patterns which have recently proved to be accurate in some respects. The occurrence of similar wing pattern elements and boundaries described by Gilbert and Nijhout in many different lineages illustrates that developmental constraint clearly plays a large role in wing pattern development. Both models can benefit from being interpreted in the context of a more current understanding of molecular genetics and development (Jiggins et al., 2017).

Both the NGP and the Gilbert model try to invoke special mechanisms to account for why some species or genera are very diverse in wing pattern whereas others are essentially monomorphic. While Gilbert asks “why are so many wing patterns available to *Heliconius*?”, I would contend that we should ask an alternate question: “why are so many wing patterns *successful* in *Heliconius*?”. Evolution of wing pattern occurs completely within the context of selection, and so the broad context of the life history of a butterfly must be considered. It is possible, for example, that some lineages are monomorphic simply because of the absence of selective pressures on wing pattern. In fact, an interplay of the presence and absence of selection alongside lineage specific adaptation is possible. The heliconiine butterflies contain interesting case studies for this: while the genera *Heliconius* and *Eueides* are relatively speciose, other genera in this group are much less diverse, including the single-species genus *Agraulis*.

Taking a broader view: Agraulis vanillae

Agraulis is a Heliconiine Nymphalid, which shared its most recent common ancestor with *Heliconius* around 30 mya (Kozak et al 2015). They are an agricultural pest in the production of passion fruit, and have a very broad distribution, ranging from the southern United States (including a recent invasion of California) through the Caribbean and into Central and South America.

Unlike *Heliconius*, *Agraulis* has a fritillary-type wing pattern; the dorsal surface of their wing is orange with black intervein elements and some black detailing along the

veins, similar to other Heliconiines (like *Dione juno*) whereas the ventral surface is orange and brown, with silver patches outlined in black. They are mainly monomorphic across their range, with some limited local variation including the presence vs absence of black markings which in some localities is sex-specific, and some size and shape variation in silver markings. They are the only member of their genus.

Like *Heliconius* and other heliconiines, larvae of *Agraulis* feed on passifloraceae and therefore acquire and incorporate cyanogenic compounds, making them distasteful to predators. Unlike *Heliconius*, *Agraulis*' orange, silver and black patterns do not closely mimic any other species, although their wing pattern is aposematic (Benson, 1971).

This species, relatively closely related to *Heliconius* but with a distinct wing pattern and facing distinct selective pressures, stands as an interesting test case for understanding how the developmental space and gene regulatory networks of the wing might evolve to affect pattern. In Gilbert's view, in which *Heliconius* has special characters that allow for pattern diversification, we might expect the developmental context of the wing gene regulatory network to be distinct and different from *Heliconius*, but in the NGP view, we might expect to find many factors in common between these three species.

Project aims

Here I address these questions of robustness versus diversity by studying the spatial and temporal patterns of gene expression in developing butterfly wings across the heliconiine phylogeny. I cut developing wings into sections that correspond to the adult postman pattern elements in order to ask what factors are differentially expressed during development. This splits the hindwing into the anterior and posterior compartment that splits the yellow bar from the rest of the hindwing of *H. melpomene* and *H. erato*, and splits the forewing into three sections along the proximodistal axis that correspond to the black proximal portion, the medial red band portion, and the distal black tip.

I characterised the expression of Wnt pathway constituents, transcription factors and other homologous differentially expressed genes in each of the three species used in this study. This has allowed me to identify transcripts which are expressed in patterns correlated with adult wing pattern.

II: MATERIALS & METHODS

Tissue sampling and dissection

Collection of larval and pupal tissue from *H. erato demophoon* and *H. melpomene rosina* was carried out as in the previous chapter. Larval forewings were removed from late 5th instar caterpillars as described, pooled left and right, and stored in the same way as hindwings. Pupae were dissected in cold PBS. Wings were removed from the pupa and cleared of peripodial membrane. The wings were then cut with microdissection scissors into 5 sections: forewing proximal, medial and distal, and hindwing anterior and posterior. The lacunae (developing veins) were used as landmarks for dissection, as indicated in Figure S4.1. These sections were stored and transported as described.

Pupal wing tissue from two pupal stages of *A. vanillae* were collected in March 2014 during the late dry season – on my later visit from July-October 2015 (during the rainy season) I was not able to acquire larval tissue from this species, and so it is excluded from the larval analysis.

RNA extraction and sequencing

This was carried out as in the previous chapter.

Mapping and annotation – H. melpomene and H. erato

Mapping and quantification in *H. melpomene* and *H. erato* was performed as in the previous chapter – analysis in these species is expedited by the use of a high quality reference genome for mapping. The LepBase genome annotation for *H. erato* v1 was generated by Alexei Papanicolaou in Van Belleghem et al, and contains 20,118 genes (Van Belleghem et al., 2017). The initial *H. melpomene* genome annotation for version 2 of the *H. melpomene* genome was generated by lifting over the version 1 annotation, with no additional reconstruction, and contains 13,178 genes (Davey et al., 2016). This annotation was improved by incorporating RNAseq datasets generated since the publication of the *Hmell* genome in 2012, including the *H. melpomene* data generated in this study. Sujai Kumar (LepBase, University of Edinburgh) used the

BRAKER1 pipeline to perform unsupervised RNAseq-based annotation (Hoff et al., 2016). GeneMark-ET and AUGUSTUS were used for iterative training and subsequent integration of RNAseq read information into the final gene predictions, generating 26,017 predicted transcripts (Lomsadze et al., 2014, Stanke and Morgenstern, 2005). Ana Pinharanda filtered these transcripts based on 90% single hit matches to repeat libraries, removing 6,532 repeat transcripts, and transferred manually curated annotations from *Hmell* and from published RNAseq studies that included curated annotations (Briscoe et al., 2013, van Schooten et al., 2016, Yu et al., 2016). This resulted in an annotation containing 20,102 genes. This newly-generated annotation is now available at LepBase.

Mapping and annotation – Agraulis vanillae

Very limited genomic resources exist for *Agraulis*; two Whole Genome Shotgun libraries generated by Kozak et al (Kozak et al., 2015) are available, as well as a DISCOVAR assembly with an N50 of 21.4 kb, in 45,022 scaffolds, and a total length of 391 Mb. There are no publicly available transcriptomic datasets from *Agraulis* other than the data generated here. As such, I took two approaches to analyzing this data; transcriptome assembly, and reference guided assembly with annotation transfer.

Transcriptome assembly - Agraulis

All paired end sequence data for *Agraulis* was assembled with the transcriptome assembler Trinity (Haas et al., 2013). This generated 87,214 contigs. Next the Trinity output was passed through TransDecoder (Haas & Papanicolaou, in prep), which annotates the transcript contigs based on the likelihood that they contain reading frames, and also based on similarity by BLAST of transcripts to reference assemblies, in this case *H. erato* and *H. melpomene*. This annotation (a GFF3 annotation of the trinity contigs) contained 24,984 genes, which compares to 20,102 annotated genes in *H. melpomene* v2 and 13,676 in *H. erato* v1.

There are two primary reasons for the high number of assembled transcripts in *Agraulis* relative to the expectation from *Heliconius*. First, the contigs are likely to include assembled non-genic material. This will include polyadenylated non-coding transcripts, partial transcripts that failed to assemble, and also likely some

environmental contamination like bacterial and viral transcripts. Although many of these contigs should be removed by TransDecoder, some may remain in the transcriptome. Second, contigs that correspond to one transcript will sometimes be separated into two transcripts; *Agraulis* has a high effective population size and thus has many polymorphisms, so some transcripts that are two alleles of the same gene may be assembled as two genes. Also, some genes will be split into two contigs, while other contigs will be fusions of two transcripts. These errors are difficult to correct in the absence of a high-quality reference genome.

Reference guided assembly and annotation transfer - Agraulis

I therefore also produced a reference guided assembly for *A. vanillae* using the DISCOVAR assembly available at Lepbase.org (Mallet et al, in prep) and the *H. erato v1 (demophoon)* genome with the program Ragout (Kolmogorov et al., 2014). Reference guided assembly uses a second (reference) genome of a higher quality to scaffold contigs. This relies on the assumption that there are no major structural rearrangements, fusions, or large insertions and deletions. The Ragout assembly stats are in Table 4.1 below; note that 48% of the DISCOVAR contigs were assembled into 84 scaffolds, with a total assembly length of 326.8 Mb, and an N50 of 9.88 Mb, which is approaching the N50 of *H. erato v1* at 10.7 Mb. We do not have a flow-cytometry estimate of the size of the *Agraulis* genome, so it is not possible to estimate accuracy of the size, and how much of the unused material should be incorporated into this assembly.

The *H. erato v1* annotation was transferred onto the *Agraulis* Ragout assembly using the program RATT (Rapid Annotation Transfer Tool). First, the *H. erato* annotation (in .gff3 format) was converted into EMBL format, using the EMBOSS package program Seqret (Rice et al., 2000). RATT takes the EMBL format annotation of *H. erato* and an alignment of the *H. erato v1* genome to the *Agraulis* Ragout assembly (generated with LASTZ), and creates an EMBL format annotation of the Ragout assembly. Seqret is then used to convert this EMBL file into a General Feature Format (GFF3) annotation of the Ragout fasta, which can be used by downstream applications. Of the 13,676 genes in *H. erato v1*, only 7,204 (or 52.7 %) were successfully transferred.

Mapping and quantification – Agraulis

Reads were aligned to the Ragout assembly using Hisat2 with varying test parameters. Generally, mapping percentage was lower than achieved with *H. melpomene* or *H. erato*. A range of parameters were tested, but default parameters gave the highest percentage of unique mapping reads. Reads were aligned to the Trinity assembly with Bowtie2, using the default parameters.

Data analysis

Statistical analysis of counts was carried out using the R package DESeq2 (Love et al., 2014) using the following generalized linear model (GLM):

$$\sim \textit{individual} + \textit{compartment}$$

In larvae, the compartments were Forewing (FW) and Hindwing (HW). In pupae, the compartments were as follows: Proximal Forewing (FP), Medial Forewing (FM), Distal Forewing (FD), Anterior Hindwing (HA), Posterior Hindwing (HPo)). Pairwise contrasts between each compartment were extracted after conducting the GLM. Additionally, samples were clustered using the DESeq2 PCA function, and the average Cook's distance value for every gene in each sample, which corresponds to the spread of variance of each sample in each gene, was also plotted. If an individual sample's average Cook's distance varies significantly from the average, it is possible that this sample is lower quality or contaminated.

Determining orthology

Orthology between differentially expressed genes in the three species was determined in two primary ways. First, a small percentage of genes are assigned homologs with other Lepidopteran genomes on LepBase, which means some genes from *H. melpomene* and *H. erato* were already assigned homologs. For the rest of the genes, as well as all genes in *Agraulis*, amino acid sequences were reciprocally searched with BLASTp, and the top hit was taken as the homolog (Altschul et al., 1997).

InterProScan results for some genes have been generated by LepBase, providing information about Orthology and functional annotation (Jones et al., 2014). This has

worked well for one-to-one homologs and gene families with distinct insect lineages. In cases where several similar copies of a gene are present, for example the Wnt ligands, genes are often assigned to the incorrect orthogroup. In some cases individual genes in complex families were manually curated.

Genes with no assigned orthogroup were compared using against the polypeptide library of *D. melanogaster* genes retrieved from FlyBase, associating them with a FBpp number and gene code based on homology with *Drosophila* genes (Gramates et al., 2017).

Table 4.1: Ragout assembly statistics	
Scaffolds:	84
Used fragments:	21,399 (47.53%)
Used fragments length:	326,791,079
Unplaced fragments:	26,218
Unplaced length:	84,876,152 (21.72%)
Introduced Ns length:	19,532,240 (5.98%)
Fragments N50:	21,413
Assembly N50:	9,877,933

Table 4.1: DISCOVAR data for *Agraulis* was assembled by reference guided assembly to the *H. erato* v1 genome. These are the assembly statistics for the read-mapped assembly.

III: RESULTS

Comparison of Agraulis annotations

The number of differentially expressed genes at each contrast was calculated for both *Agraulis* annotations (Table 4.2). At day 1, the Ragout/RATT annotation had 882 genes differentially expressed, just 49% of the 1780 genes scored as differentially expressed in the Trinity/TransDecoder transcriptome. At Day 2, this was 30%. This difference mirrors the difference in total number of genes in each version (Table 4.1).

Examination of differential expression of the transcription factor *Ubx* in the two annotations illustrates the relative costs and benefits of each (Figure 4.1). The Trinity/TransDecoder transcriptome maintains the *H. erato* gene ID, and while the gene was significantly differentially expressed under both annotations, both the total number of reads mapping to the transcript and the fold-change are higher with the Ragout/RATT genome annotation than with the Trinity/TransDecoder transcriptome. The *Ubx* gene in the *H. erato v1* annotation spans 155,207 bp and is a multiexonic transcript including 564 bp of coding sequence plus noncoding exons and 5' and 3' UTRs totalling 6595 bp. This annotation transferred onto the Ragout *H. erato* genome with high fidelity. The Trinity scaffold for *Ubx* is 4885 bp in length (note that by its nature, this does not include intronic sequence). This includes 588 bp of coding sequence (indicating 8 additional amino acid residues relative to the Ragout annotation) and 4297 bp of non-coding sequence. There is a discrepancy of 2298 bp between the two annotations, which likely accounts for the difference in read mapping.

However, in spite of this difference in read mapping and counting between the two annotations, I proceeded with the Trinity/TransDecoder transcriptome for two reasons. Firstly, as a higher number of genes were identified, it is less likely that there are important genes missing from this annotation, and I was able to identify many homologs despite the lack of positional homology. Second, the average percentage of unmapped reads was 20.43% to the Ragout genome vs 8.51% to the Trinity transcriptome, indicating that many more reads could be counted using this method, increasing the power of the analysis (Table S4.1) (Figure 4.1)

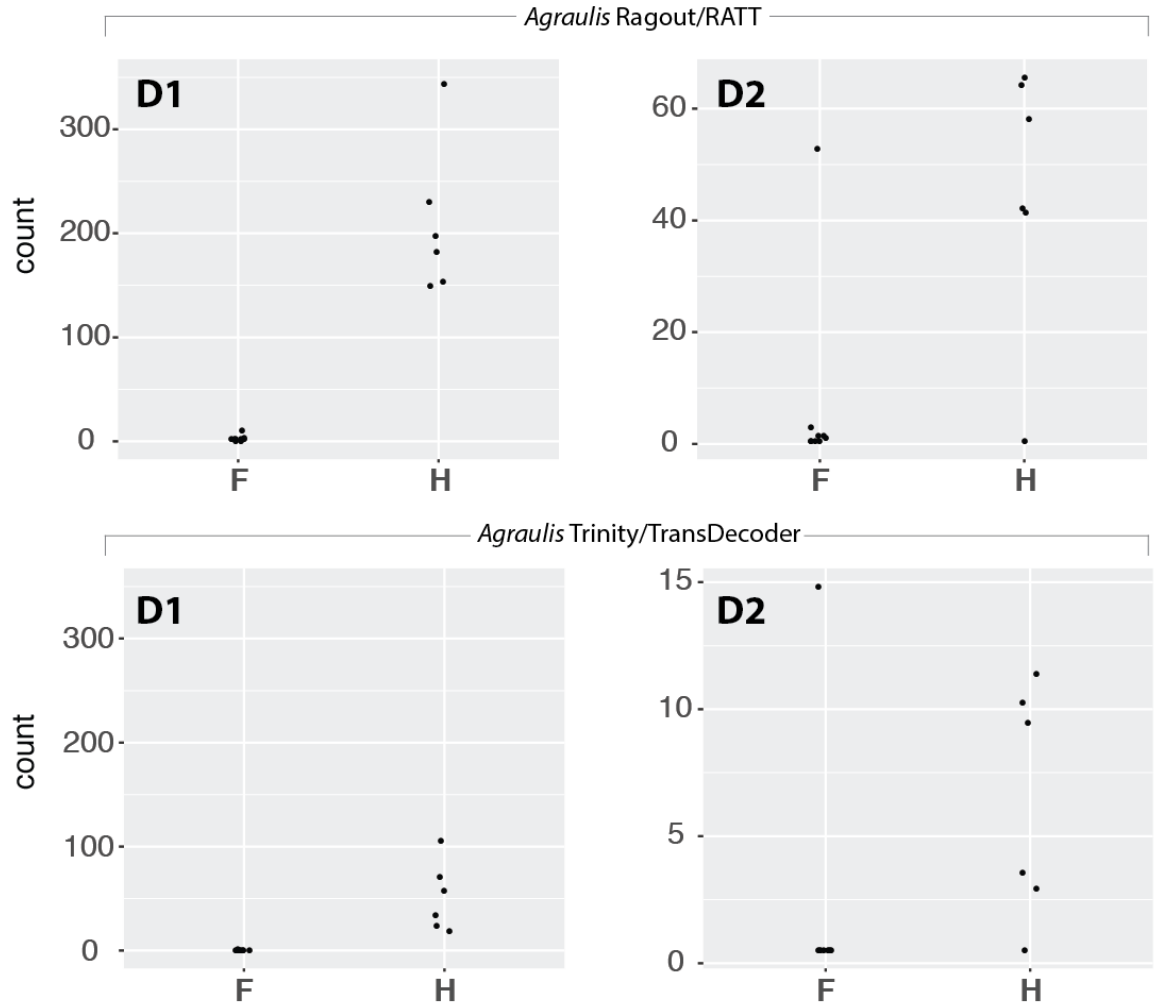


Figure 4.1: A comparison of the counts mapping to the annotation for the gene *Ubx* in the two *Agraulis* genome annotation vs the transcriptome assembly. F-forewing, H-hindwing. Note that the transcriptome assembly counts fewer reads as mapping to the gene *Ubx*.

Table 4.2

DAY 1	<i>Agraulis</i> Trinity / TransDecoder assembly		<i>Agraulis</i> Ragout / RATT annotation	
	+	-	+	-
FP vs FM	308	56	121	22
FM vs FD	9	4	0	0
FP vs FD	405	187	212	130
HA vs HP	137	24	62	12
FP vs HA	149	59	61	32
FP vs HP	491	231	254	153
FM vs HA	233	416	98	189
FM vs HP	152	59	64	29
FD vs HA	374	520	184	256
FD vs HP	259	99	111	58
TOTAL	1780		882	
DAY 2	<i>Agraulis</i> Trinity / TransDecoder assembly		<i>Agraulis</i> Ragout / RATT annotation	
	+	-	+	+
FP vs FM	12	14	0	0
FM vs FD	37	24	6	2
FP vs FD	7	12	0	0
HA vs HP	4	0	3	0
FP vs HA	1	2	0	3
FP vs HP	2	0	1	0
FM vs HA	22	23	7	12
FM vs HP	37	13	8	2
FD vs HA	11	10	2	2
FD vs HP	38	5	10	0
TOTAL	167		40	

Table 4.2: Numbers of differentially expressed gene in each contrast for the Trinity/TrnasDecoder transcriptome assembly and the Ragout/RATT genome annotation at both day 1 and day 2.

Data counts and clustering

PCA analysis showed that at Day 1, sample clustering by compartment is clear (Figure 4.2). In *Agraulis* and *H. melpomene*, distinct clusters for forewing and hindwing are also present. However, at Day 2, clustering by compartment is not evident in any species, and some clustering by individual can be seen. Genes with an adjusted P-value of less than 0.1 were considered to be significantly differentially expressed.

Average Cook's distance for each sample at each stage did not indicate that any individual samples were skewed relative to other samples, (Figure 4.3). The average percentage of reads per sample that do not map is 11.8% (supplementary table 1), compared to a previous RNAseq study in *H. melpomene* in which 50.42% of reads did not map (Walters et al., 2015). The percentage of reads failing to map in *H. erato* day 2 samples is higher than others at 18.9%. The *H. melpomene* day 1 anterior hindwing sample 24D has 32.7% of reads failing to map. These differences in sample quality do not account for clustering by individual in all species at day 2.

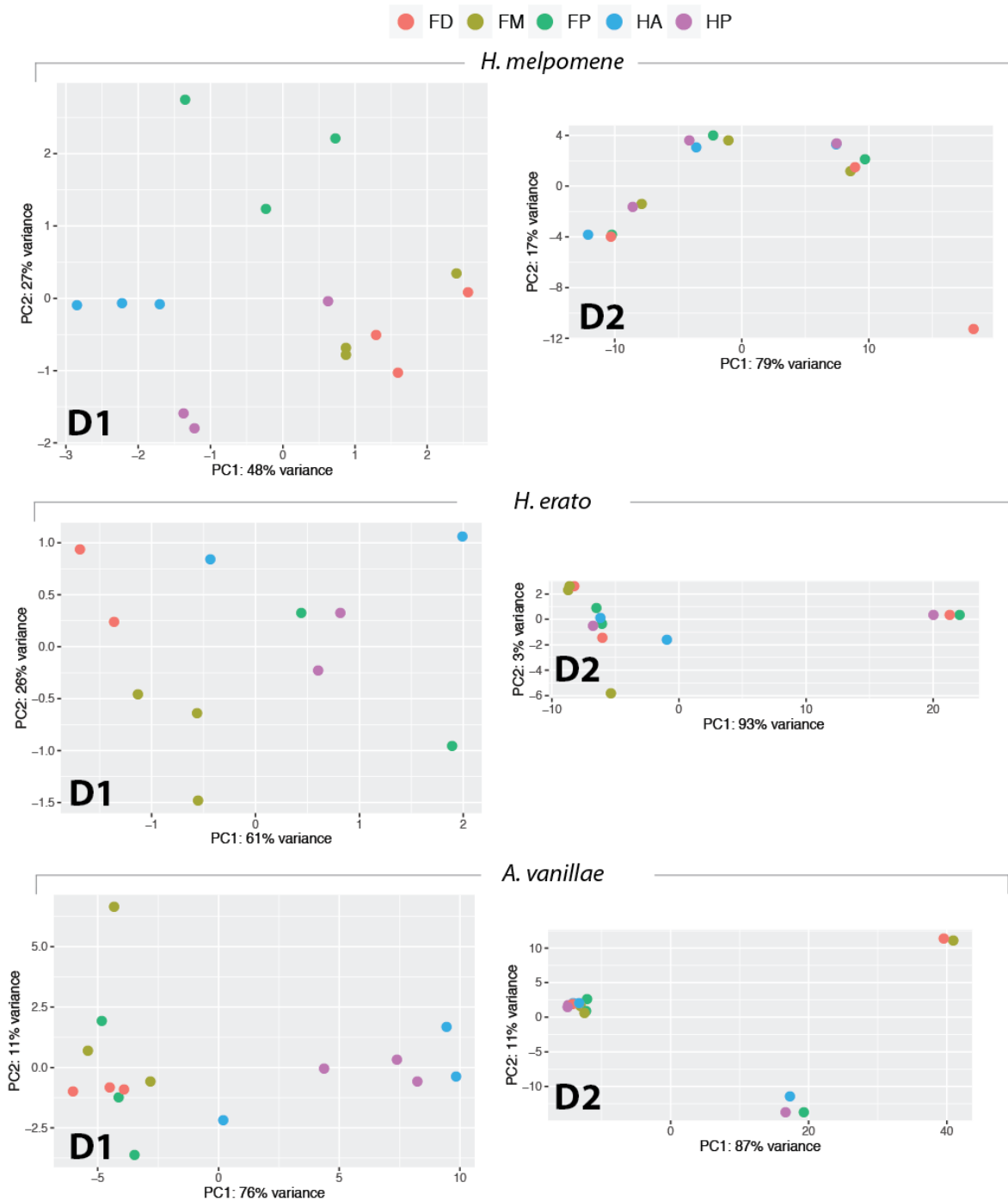


Figure 4.2: Principle component analysis of RNAseq data used in this chapter. Each point corresponds to a sample, with colours corresponding to wing compartment. (FP – proximal forewing, FM – medial forewing, FD – distal forewing, HA – anterior forewing, HP – posterior forewing).

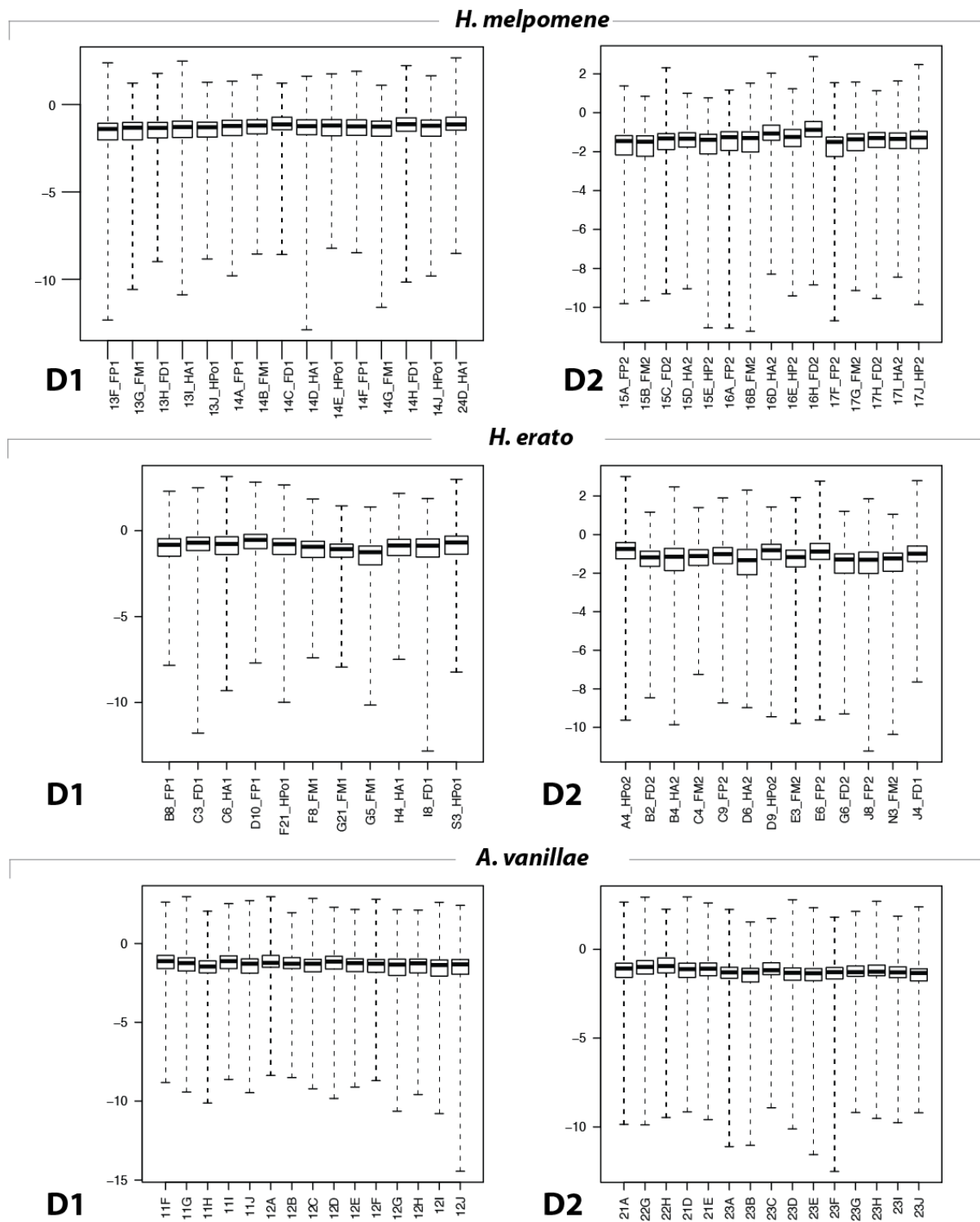


Figure 4.3: Mean Cook's distances for each individual. The variance is calculated for each gene in the data set, and a Cook's distance calculated for every individual at that gene. A high value indicates high deviation from the mean, and so an elevated average may indicate a problem with an individual sample. Here, no individual sample(s) deviated from the others in their cohort.

Differential expression in larvae

In all of the analyses I will focus specifically on TFs and the Wnt signaling pathway, as these are known to play an important role in wing specification.

In all, 209 genes are differentially expressed between *H. melpomene* larval forewings and hindwings, versus 77 in *H. erato*. In total, 28 of these genes are differentially expressed in both species (Table 4.3). This includes the transcription factor *Ubx*, the notch pathway repressor *pigs*, and 9 genes with no homology to known transcripts (Table 4.4).

In *H. melpomene* larvae, the TFs *Ultrabithorax*, *mirror* and *cubitus interruptus* are more highly expressed in larval hindwing than forewings, whereas in *H. erato*, only the TF *Ubx* is differentially expressed in larval wings. In *H. melpomene*, the Wnt pathway genes *multiple wing hairs (mwh)*, and *apolipophorin (Rfabg)* are differentially expressed, whereas no Wnt pathway genes are differentially expressed in *H. erato* (Table 4.4). Of the 9 unannotated genes that are differentially expressed in both species, 7 are upregulated in forewings.

LARVAE	<i>H. melpomene</i>		<i>H. erato</i>	
	+	-	+	-
Forewing vs Hindwing	146	63	15	62
Genes differentially expressed in both species: 28				

Table 4.3: Number of genes differentially expressed between larval forewings and hindwings in *H. erato* and *H. melpomene*. Larval wings of *Agraulis* were not sequenced.

Genes upregulated in forewings		
<i>nov_gene_002001</i>		
<i>ham</i>	hamlet	neuron fate selection, type II neuroblast maturation
<i>AOX3</i>	Aldehyde oxidase 3	
<i>Dhpr</i>	Dihydropteridine reductase	
<i>nov_gene007136</i>		
<i>PK2-R1</i>	Pyrokinin 2 receptor 1	GPCR, binds Hug (an NT and chitin synthesis regulator)
<i>nov_gene009808</i>		
<i>CG10298</i>	no annotated func	
<i>nov_gene034094</i>		
<i>nov_gene034095</i>		
<i>Pepck</i>	Phosphoenolpyruvate carboxykinase	
<i>nov_gene035885</i>		
<i>nov_gene036476</i>		
<i>Nrk</i>	Neurospecific receptor kinase	axon pathfinding, rhabdomeere elongation
<i>nov_gene045731</i>		
Genes upregulated in hindwings		
<i>nrv2</i>	nervana 2	Na/K ATPase subunit 2.
<i>Cht2</i>	Chitinase 2	
<i>AdamTS-A</i>	ADAM metalloproteinase with thrombospondin type 1 motif A	secreted matrix metalloprotease (mutants show apical surface irregularities).
<i>CG3168</i>		anion transmembrane transporter
<i>Ubx</i>	Ultrabithorax	homeodomain transcription factor
<i>pigs</i>	pickled eggs	microtubule binding, negative regulator of notch
Genes differentially regulated in both species, but in the opposite direction		
<i>r-l</i>	rudimentary-like	orotate phosphoribosyltransferase
<i>nov_gene013098</i>		

Table 4.4: Gene codes and functions of homologous genes which are differentially expressed in both *H. melpomene* and *H. erato*.

DAY 1	<i>H. melpomene</i>		<i>H. erato</i>	
	+	-	+	-
FP vs FM	207	79	45	6
FM vs FD	321	1	0	0
FP vs FD	386	177	772	268
HA vs HP	481	264	22	74
FP vs HA	434	348	588	125
FP vs HP	396	344	392	98
FM vs HA	662	1020	73	158
FM vs HP	139	115	35	174
FD vs HA	848	1224	34	101
FD vs HP	244	172	43	199
TOTAL	2848		1713	

Table 4.5: Differential expression of genes in each contrast in *H. melpomene* and *H. erato* at pupal day 1.

DAY 2	<i>H. melpomene</i>		<i>H. erato</i>	
	+	-	+	-
FP vs FM	0	0	1001	202
FM vs FD	0	1	1	5
FP vs FD	0	1	1477	768
HA vs HP	0	0	111	21
FP vs HA	3	1	879	181
FP vs HP	1	1	107	21
FM vs HA	3	2	3	157
FM vs HP	1	1	0	2
FD vs HA	182	107	3	77
FD vs HP	61	32	11	63
TOTAL	319		2663	

Table 4.6: Differential expression of genes in each contrast in *H. melpomene* and *H. erato* at pupal day 1.

DAY 1 Overlaps	Differentially expressed in all three	Pairwise vs <i>Agraulis</i>		Pairwise vs <i>H.</i> <i>melpomene</i>		Pairwise vs <i>H.</i> <i>erato</i>	
		<i>H. e</i>	<i>H. m</i>	<i>H. e</i>	<i>A. v</i>	<i>H. m</i>	<i>A. v</i>
FD_HA	125	59	185	125	20	151	76
FD_HP	30	35	31	45	3	68	54
FM_FD	0	0	0	0	0	0	0
FM_HA	137	82	125	189	11	237	105
FM_HP	12	17	14	37	3	53	21
FP_FD	72	95	57	126	4	159	135
FP_FM	21	15	26	25	2	30	19
FP_HA	34	38	27	109	5	136	45
FP_HP	26	42	82	52	9	66	49
HA_HP	4	4	22	35	1	53	4
TOTAL	617						

Table 4.7: Numbers of genes that were differentially expressed in all three of *H. melpomene* (*H. m*), *H. erato* (*H. e*) and *Agraulis* (*A. v*) at day 1. Also listed are the numbers of homologous genes differentially expressed in two of the three species.

DAY 2 Overlaps	Differentially expressed in all three	Pairwise vs <i>Agraulis</i>		Pairwise vs <i>H.</i> <i>melpomene</i>		Pairwise vs <i>H.</i> <i>erato</i>	
		<i>H. e</i>	<i>H. m</i>	<i>H. e</i>	<i>A. v</i>	<i>H. m</i>	<i>A. v</i>
FD_HA	0	1	2	1	5	2	1
FD_HP		1	5	1	2	5	1
FM_FD		0	0	0	0	0	0
FM_HA		4	0	4	0	0	4
FM_HP		0	0	0	0	0	0
FP_FD		3	0	2	0	0	3
FP_FM		4	0	4	0	0	4
FP_HA		1	2	1	0	2	1
FP_HP		1	0	1	0	0	1
HA_HP		3	0	3	0	0	3
TOTAL							

Table 4.8: No genes were differentially expressed in all three of *H. melpomene*, *H. erato* and *Agraulis* at day 2. Also listed are the numbers of homologous genes differentially expressed in two of the three species.

Differential expression and homologous genes

Tables 4.5 and 4.6 list the number of genes which are differentially expressed in *H. melpomene* and *H. erato*, at each stage in each contrast. Numbers of differentially expressed genes for *Agraulis* are listed in Table 4.2. Similar numbers of genes are differentially expressed at day 1 in the three species, but at day 2, both *H. melpomene* and *Agraulis* have many fewer differentially expressed genes than *H. erato*, reflecting the lack of compartment-clustering in the PCAs at day 2.

Tables 4.7 and 4.8 list the numbers of genes that are differentially expressed in more than one species. At Day 1, a high proportion of genes which are differentially expressed in each species are also differentially expressed in at least one other species. In contrast to this, at day 2 no genes are differentially expressed in all three species and very few genes are differentially expressed in two species – this is a reflection of the low levels of differential expression in both *Agraulis* and *H. melpomene*.

Genes were then clustered into groups that show similar patterns of expression (Figure 4.4). Group A is high in the proximal forewing and anterior hindwing, but low everywhere else; Group B and C show more gradually increasing and decreasing gradients of expression across the forewing; group D is high in the medial forewing and group F is low in the medial forewing; and group E is low in the forewing and high in the hindwing. Some groups are enriched in some gene types and species. For example, *H. melpomene* is overrepresented with type A and underrepresented with type D.

Particular focus has been given to constituents of the Wnt signalling pathway due to the known role of WntA and other Wnts in wing patterning in *Heliconius* and other butterflies, and also to the transcription factors due to their role in providing patterning information in the context of the developing wing.

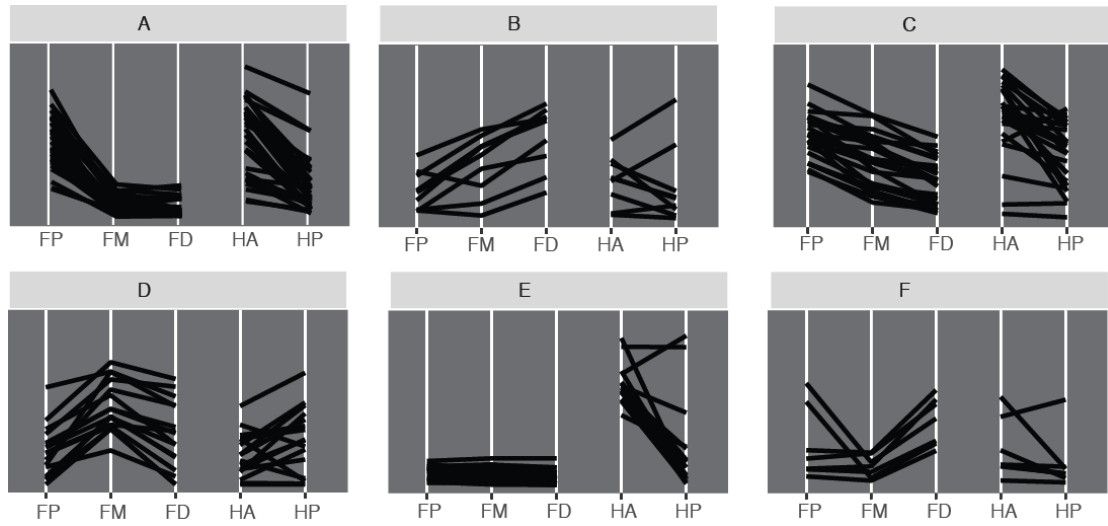


Figure 4.4: 6 prominent groups of gene expression profile were observed within the data set; these groups were arbitrarily given the names A to F.

Expression profile of Wnt pathway constituents is varied between species

All of the 12 Wnt pathway constituents which are differentially expressed in the forewings of *H. erato* at day 1 (including *WntA*), and most of the 18 at day 2, are expressed in a very similar pattern – high in FP and low in FM and FD (i.e. group A), though at day 2, *Wnt6* and *Wnt5b* also have an increased expression in FD) (Figure 4.5 and 4.6). This expression profile of pathway constituents can be correlated to the expression of *WntA* in late larval wing discs, recorded by Martin et al (2012). This pattern includes many intracellular pathway components like *Axin*, *B-catenin*, *dishevelled*, *Gilgamesh* and others. This could be a consequence of a positive feedback loop into Wnt signalling – the presence of *WntA* at higher concentrations could lead to a reinforcement of Wnt pathway constituents in this region of the wing.

A number of factors are also differentially expressed in *H. erato* Day1 hindwings, notably *WntA* itself as well as *Wnt5* and *stan* (*starry night*), which is critical for establishing planar cell polarity (PCP). Three Wnt genes are differentially regulated in the hindwings of *H. erato* at day 2, these are *Wnt6* and *Wnt5b*, which are higher in HA, and the serine-threonine kinase *dco* (*discs overgrown*), which is higher in HP.

In stark contrast to the regularity of differential expression in *H. erato*, the Wnt pathway constituents in *H. melpomene* are expressed in a variety of different patterns (Figure 4.7). Some, like *WntA* itself and the PCP-regulator *Kermit* are upregulated in FM, some are high in FP and low in FD such as *Wnt6*, *Wg*, *notum* and *basket*, whereas others are high in FD and low in FP, such as *Wnt2* and *mwh*.

In *H. melpomene* hindwings, two genes are upregulated in the anterior; the receptor *fz2* (*frizzled-2*, the main Wnt-pathway receptor) and the extracellular matrix enzyme *Notum*. *Wnt6*, *Rho1*, *Ext2*, *Wg*, *mwh*, *VhaM8.9* and *fz3* are all upregulated in the posterior compartment. Particular note should be paid to *fz2* and *fz3*, both Wnt receptors but expressed in opposition to each other. In *H. melpomene* at day 2, there are no differentially expressed Wnt pathway components.

Agraulis (Figure 4.8 and 4.9) has fewer Wnt genes differentially expressed in the forewing – at day 1, *Wnt5b* and *wntless* (a transmembrane factor required for Wnt ligand secretion) is high in FP, and *stan* (*starry night*) is high in FM. At day 2,

armadillo (β -catenin) and *mwh* (*multiple wing hairs* – an actin-binding component of the Wnt-PCP pathway) are the only differentially expressed Wnt pathway genes, and are high in FM and low elsewhere.

Differential expression between the anterior and posterior compartment of the hindwing of *Agraulis* may clarify the differing effect of *WntA* manipulation in each area. At day 1, *wntless*, *CG8786* and *Wnt5b* are upregulated in the anterior hindwing whereas *mwh* is upregulated in the posterior hindwing. There are six annotated copies of the gene *Rfabg* which are differentially expressed. This gene is also known as apolipophorin, a constituent of the major haemolymph lipoprotein lipophorin, which can carry lipophilic hormones and signalling proteins. Different copies of *Rfabg* have a variety of differential expression profiles.

Figure 4.5 DE in *H. mel* DE in *Agraulis*

H. erato Wnt pathway, Day 1

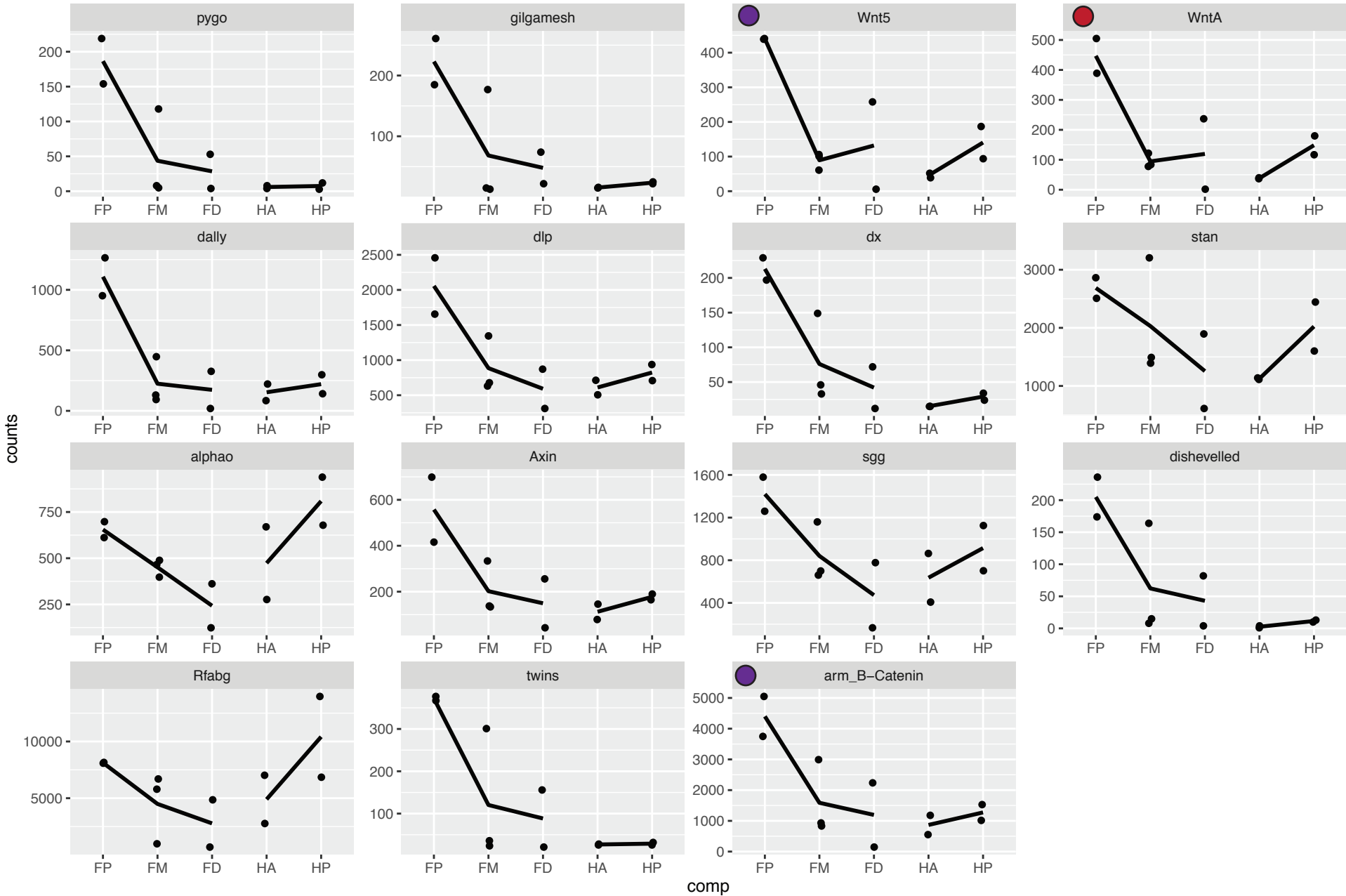
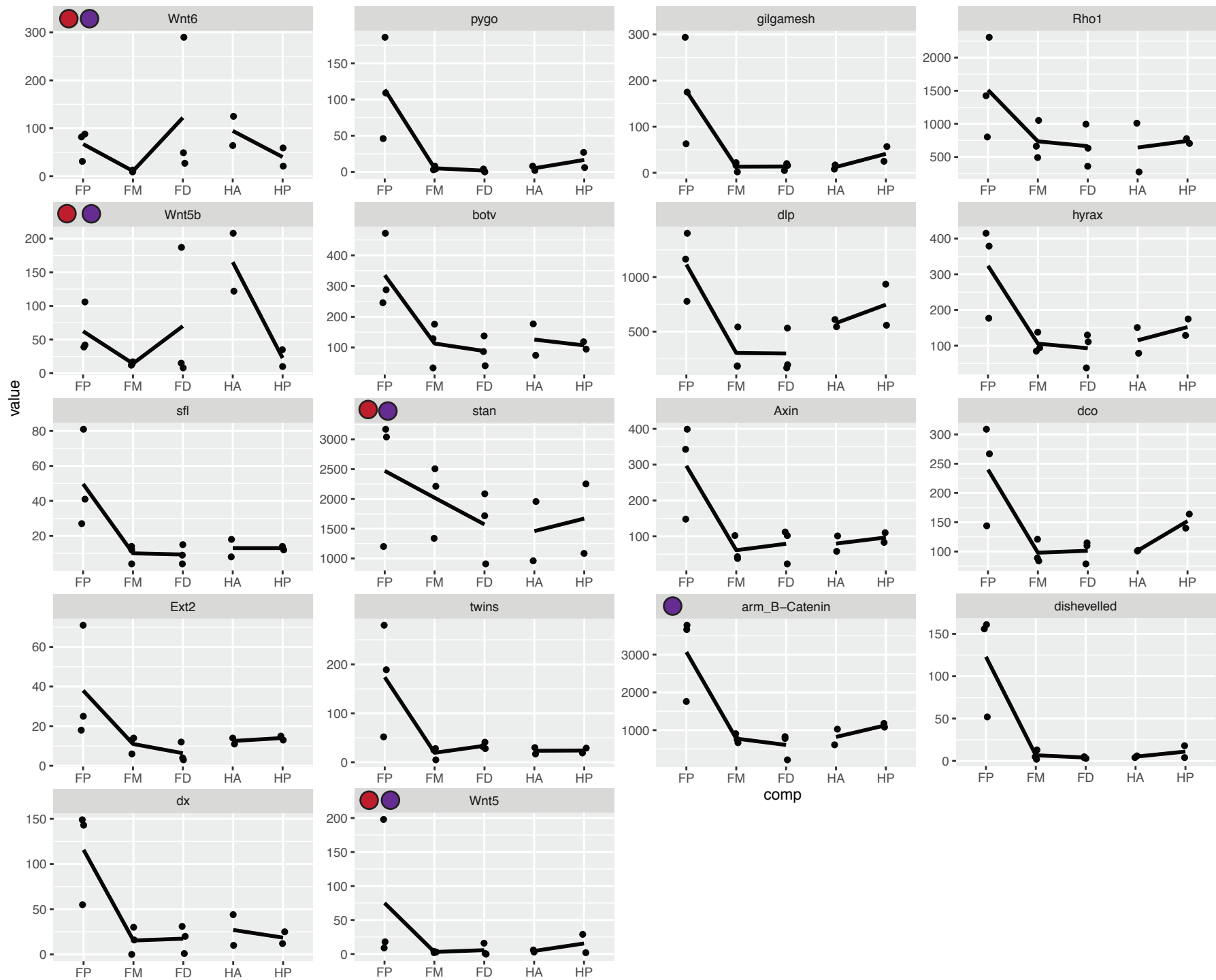


Figure 4.6

● DE in *H. mel* ● DE in *Agraulis*

H. erato Wnt pathway, Day 2





● DE in *H. mel* ● DE in *H. erato*

Agraulis Wnt pathway, Day 1

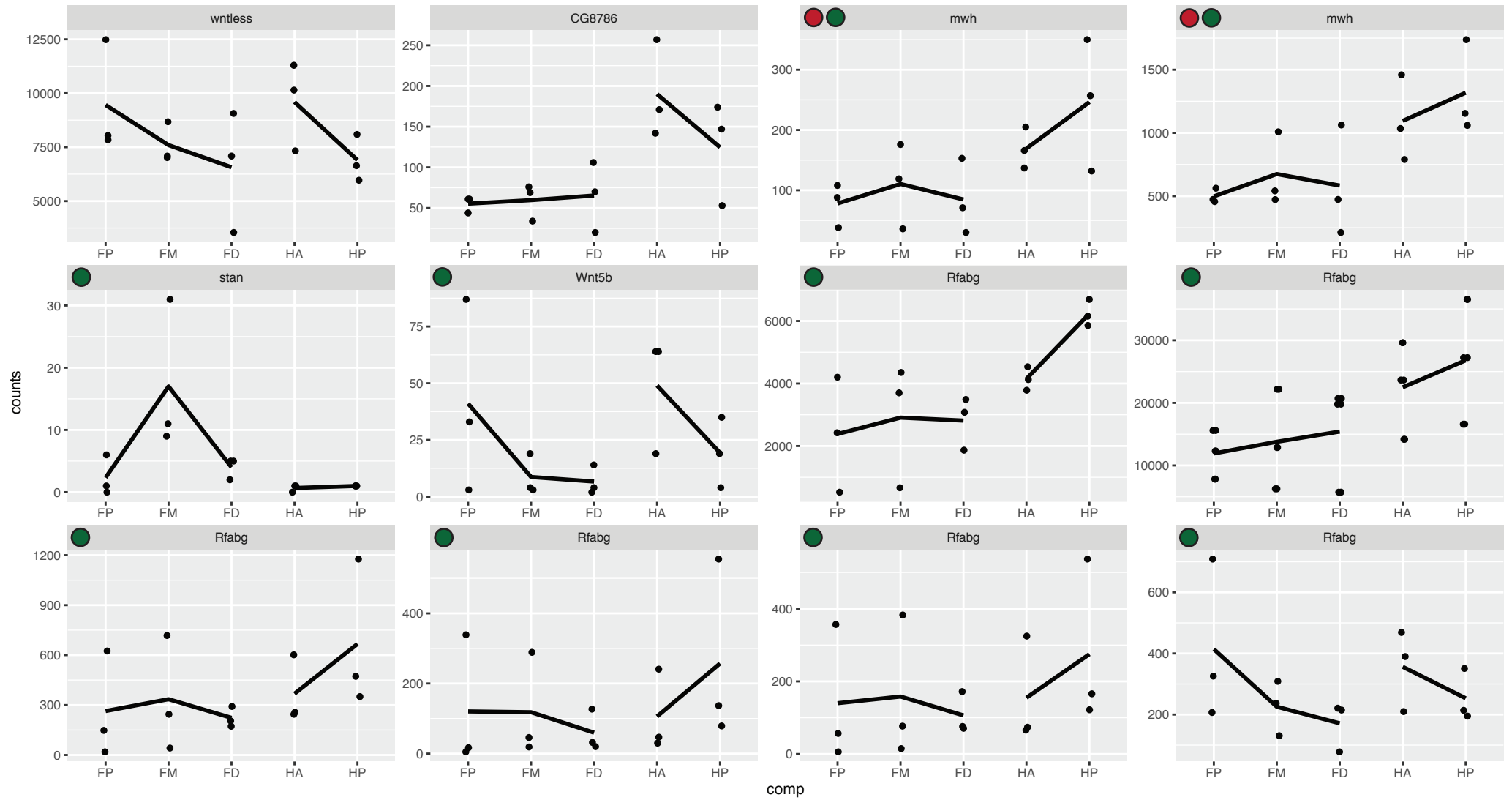
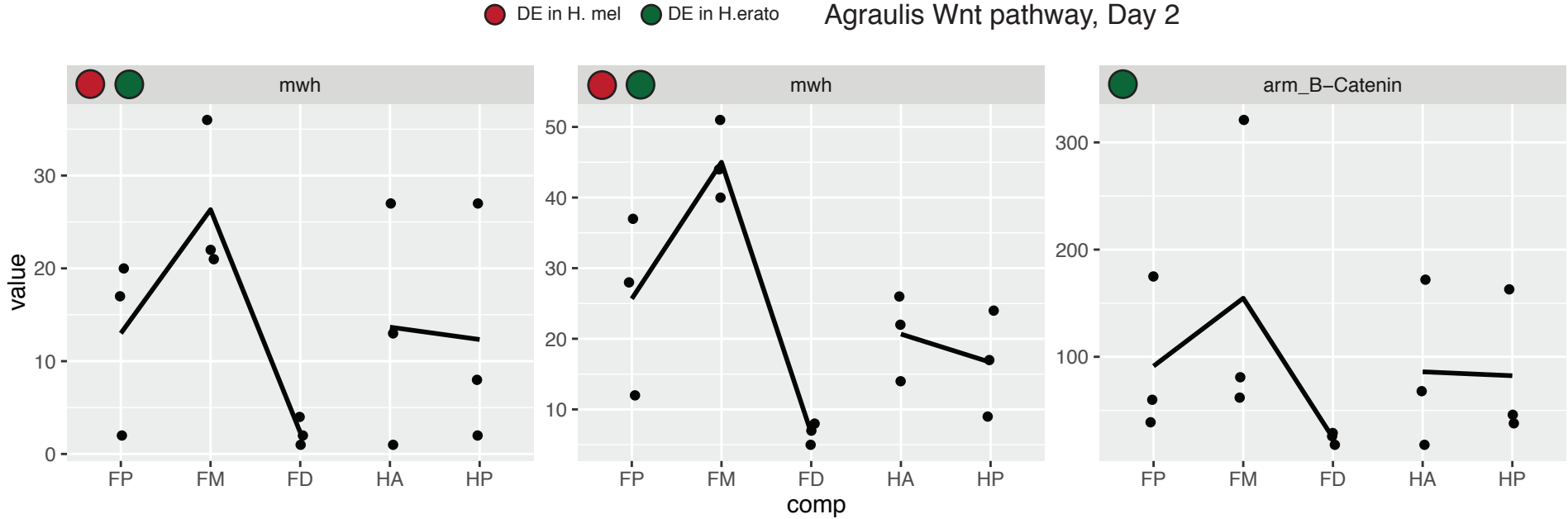


Figure 4.9



Transcription factors

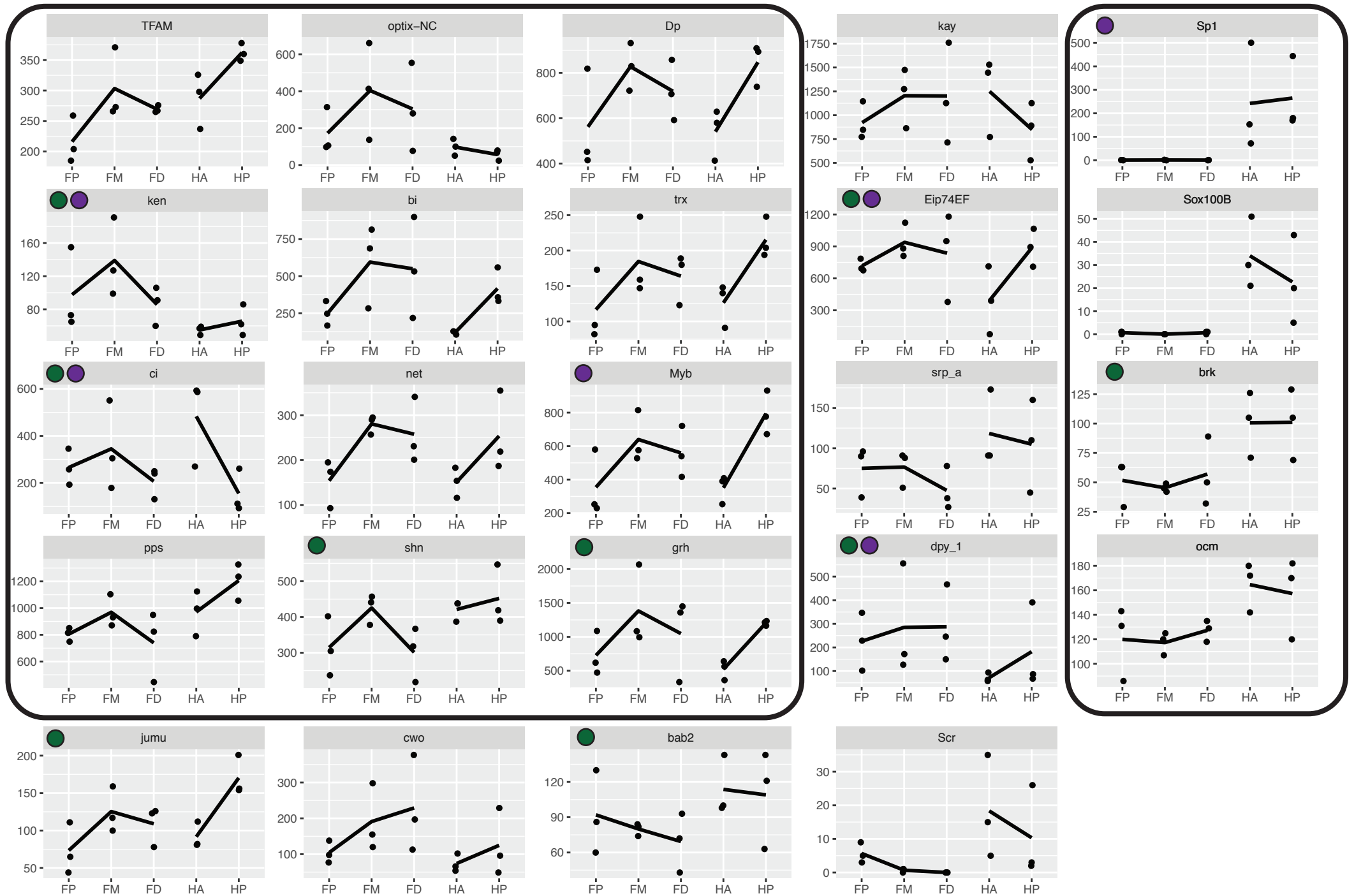
A large number of transcription factors are differentially expressed across both the fore- and hindwing of developing pupae (Figures 4.10-4.19). A number of these transcription factors match their known expression profiles either from immunohistochemistry of *Heliconius* wings or by analogy with gene expression in *Drosophila* wings. These include *Ultrabithorax* (*Ubx*), expressed only in the hindwing, *homothorax* (*hth*), which is expressed only in the proximal forewing and anterior hindwing, *distal-less* (*dll*), which is expressed in an increasing gradient from proximal to distal, *mirror* (*mirr*), which is expressed in the proximal forewing and anterior hindwing, and *cubitus interruptus* (*ci*), which is expressed in the anterior compartment of the hindwing. The recapitulation of these expression profiles serves as validation that the experimental design is capable of detecting differential expression of transcription factors in developing wings.

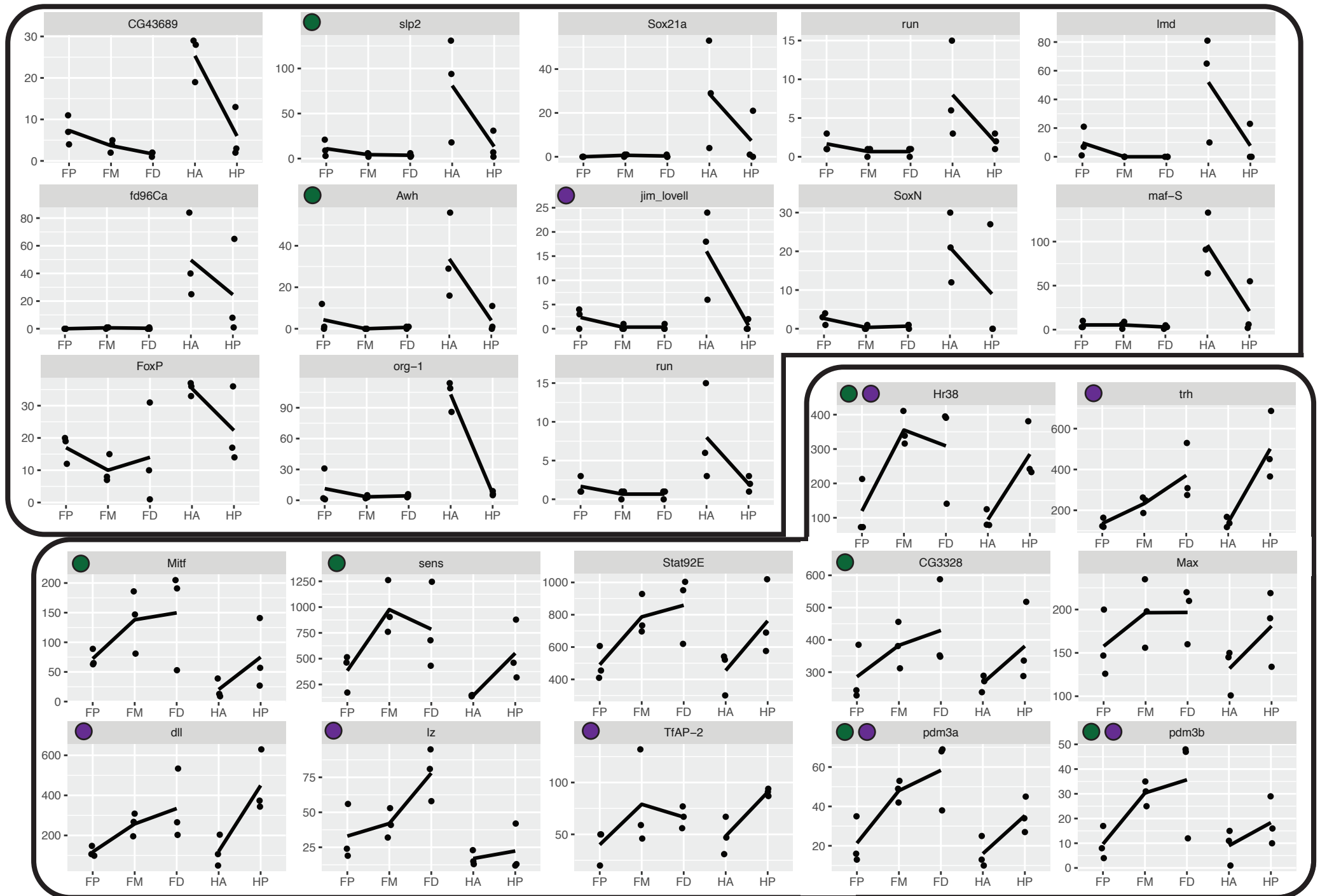
Several of the transcription factors that are differentially expressed in all three species are associated with development of imaginal discs – either wing discs, imaginal discs generally, or specifically with other imaginal discs, in particular related to the eye and the genitals. For example *Arrowhead* (*Awh*, eye disc development), *bunched* (*bun*, eye development), *lozenge* (*lz*, compound eye development and genital morphogenesis), *ken* and *barbie* (*ken*, genital morphogenesis). In *Agraulis*, *prospero*, *eyeless* and *shnurri* (*shn*, wing vein morphogenesis) are also differentially expressed.

Additionally, many of the transcription factors are linked to neurogenesis and the nervous system. *Senseless* (*sens*), *pdm3* (an olfaction TF), *glial cells missing* (*gcm* – involved in differentiation of lateral glial cells and specific neurons), *nervy* (*nvy*, axon guidance and chetae morphogenesis). Other factors have specific associations with cuticle or bristle development such as *nvy*, *grainy-head* (*grh*), as well as multiple copies of *dumpy* (*dpy*).

Some transcription factors linked to signaling pathways are consistently differentially expressed. Both notch pathway TF *cubitus interruptus* (*ci*) and the JNK pathway TF *jun-related antigen* (*jra*) are consistently differentially expressed in all three species, with *ci* upregulated in the anterior hindwing and *jra* upregulated in the posterior hindwing. The transcription factor *bric a brac 2* (*bab2*) is differentially expressed –

this gene is part of a proximal-distal gene regulatory module, and is also linked to abdominal pigmentation pattern in *Drosophila*. The TF *bunched* (*bun*), which is activated in the Dpp signal cascade and which activates notch, is expressed in a decreasing gradient from proximal to distal in forewings of *H. melpomene* and *H. erato*, but is expressed in an increasing gradient in *Agraulis*.





● DE in *H. erato* ● DE in *Agraulis*

H. melpomene TFs Day 1 page 3

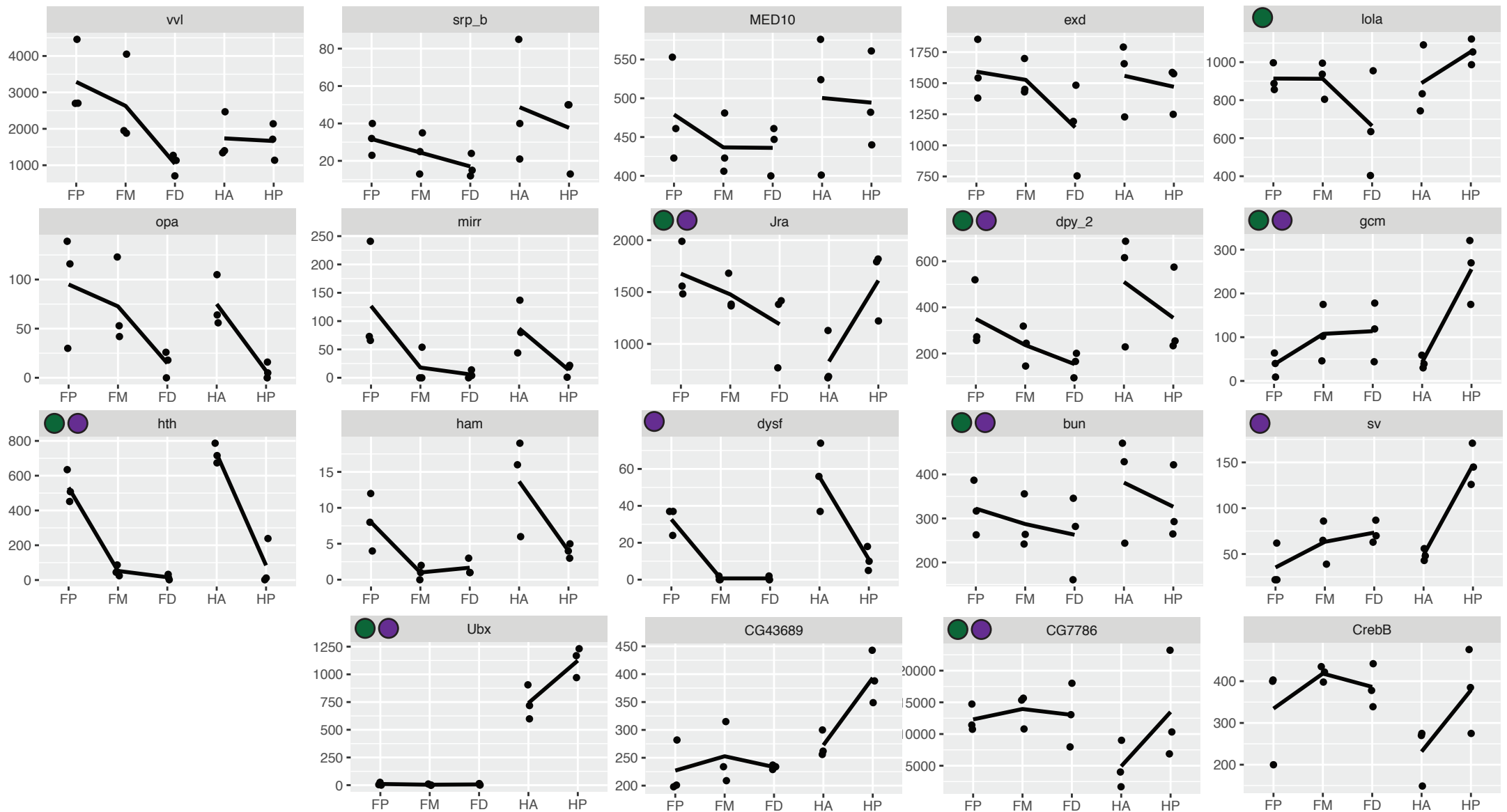
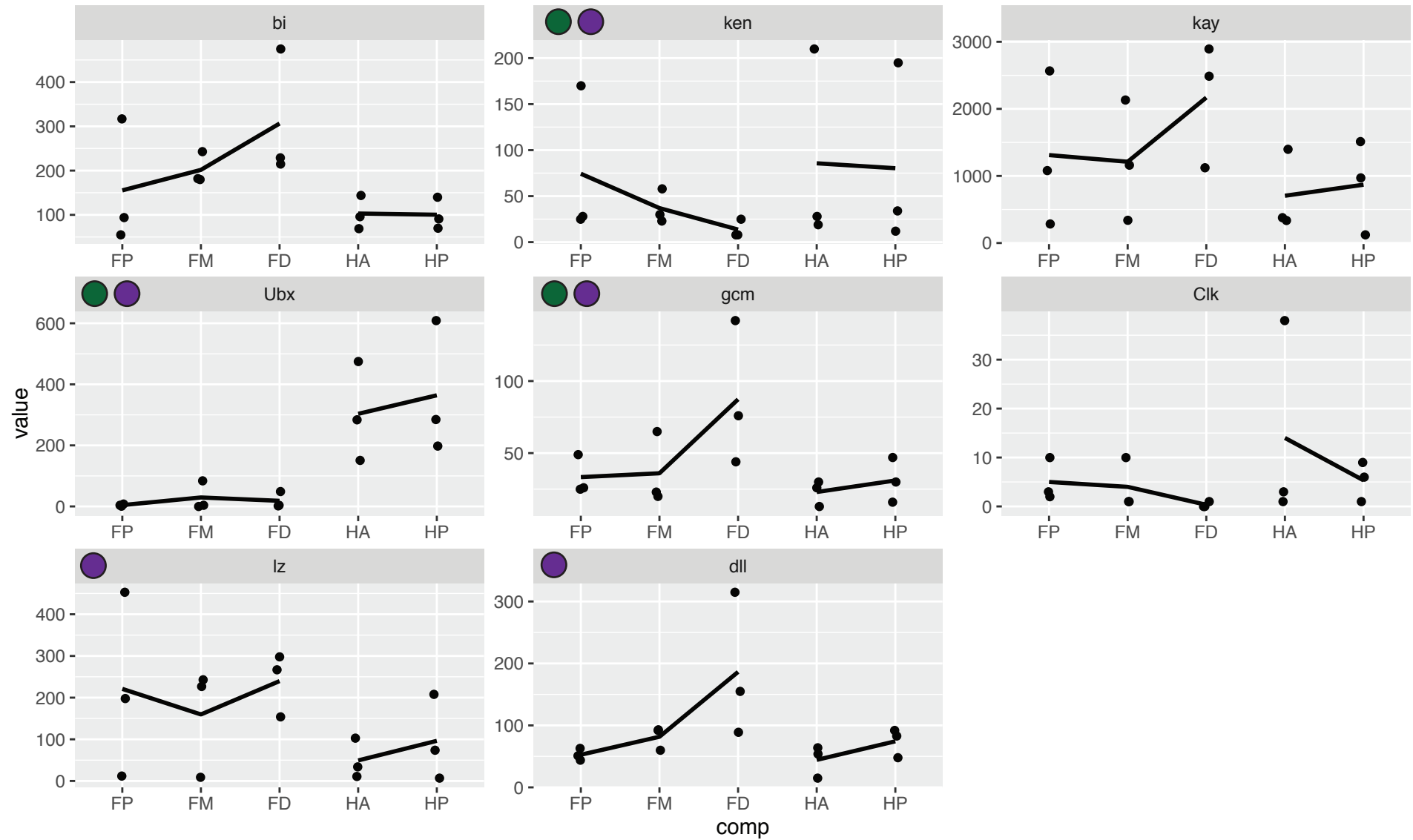
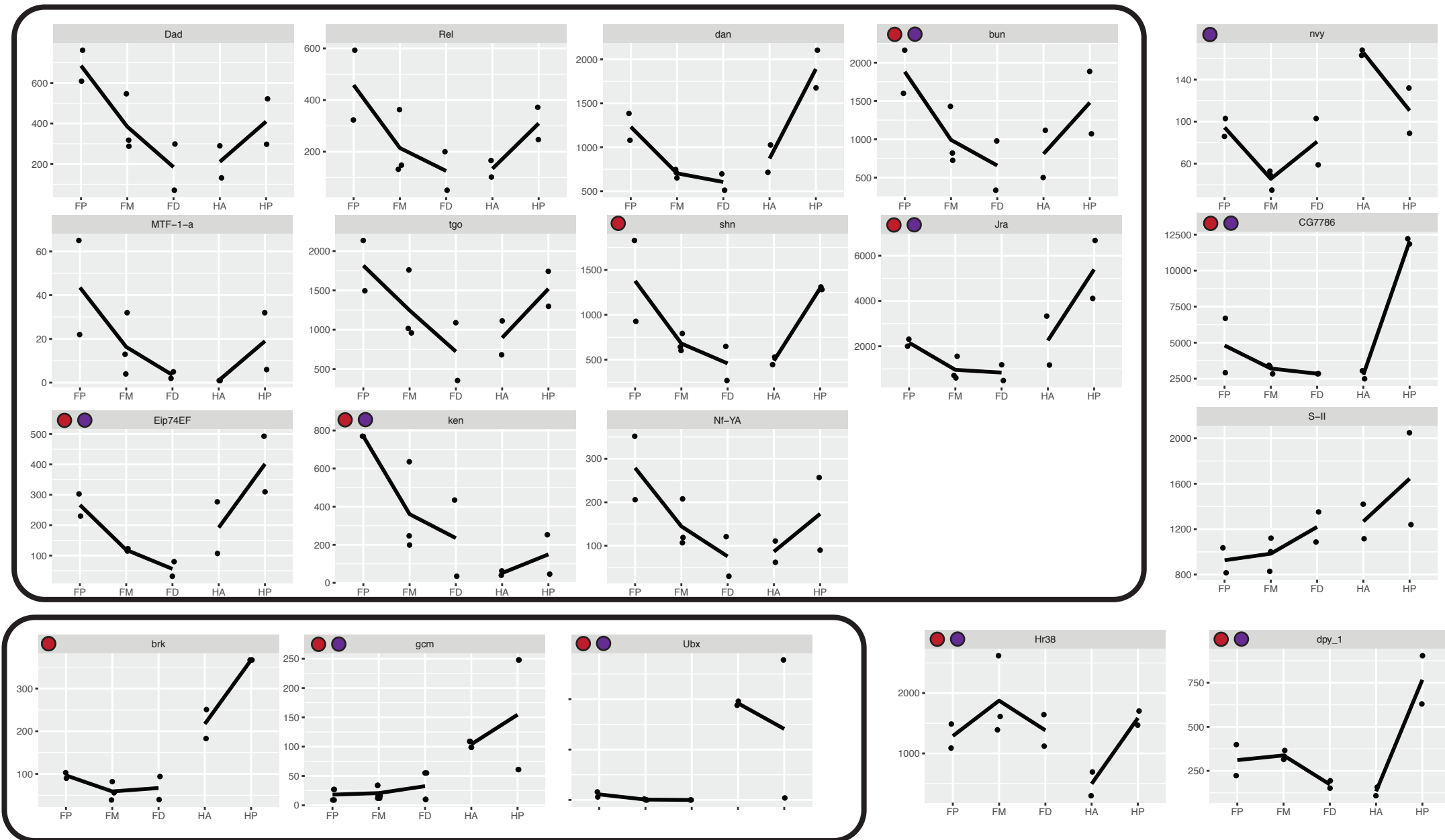


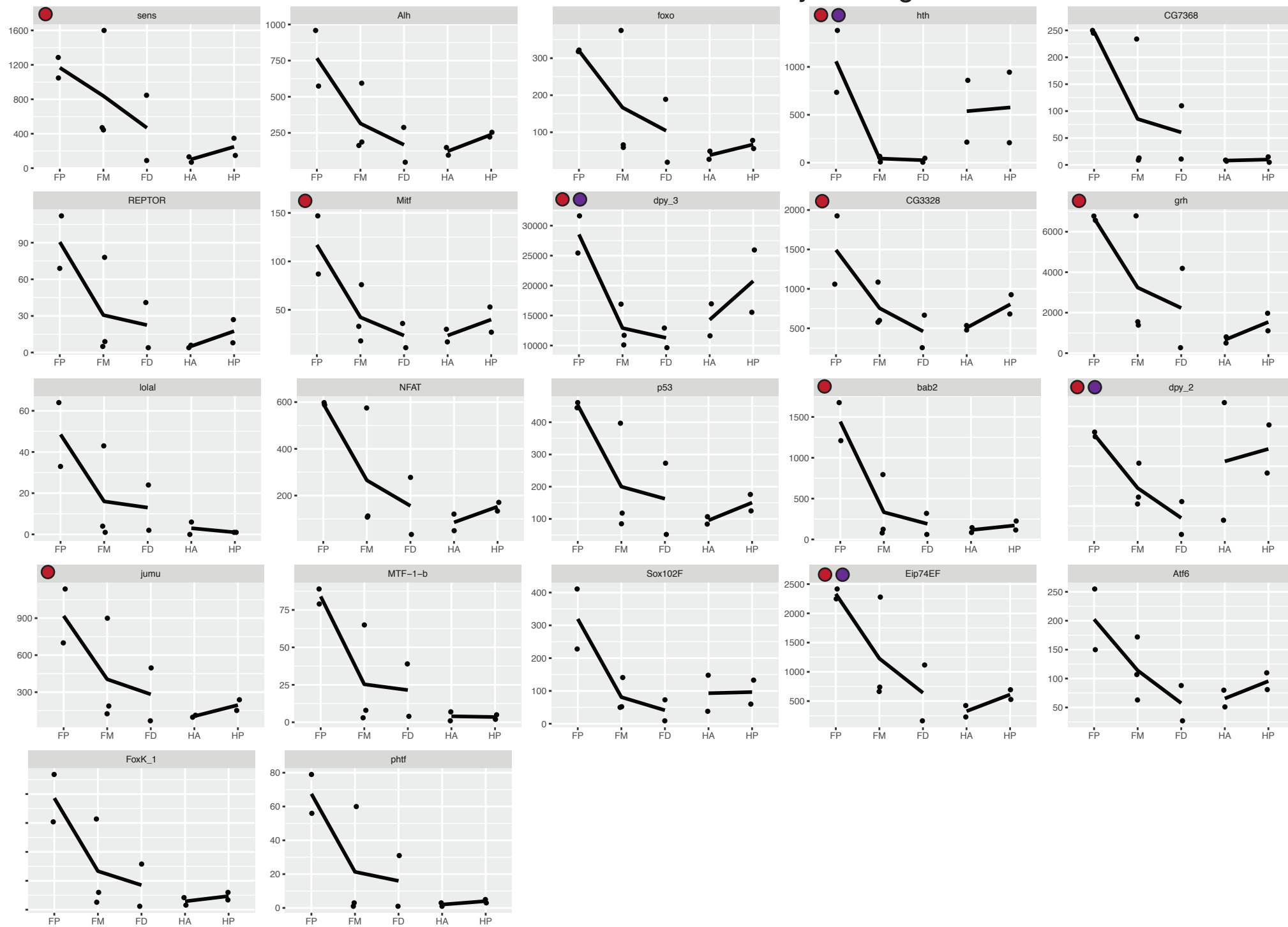
Figure 4.13

● DE in *H. erato*
● DE in *Agraulis*
 H melpomene TFs, Day 2

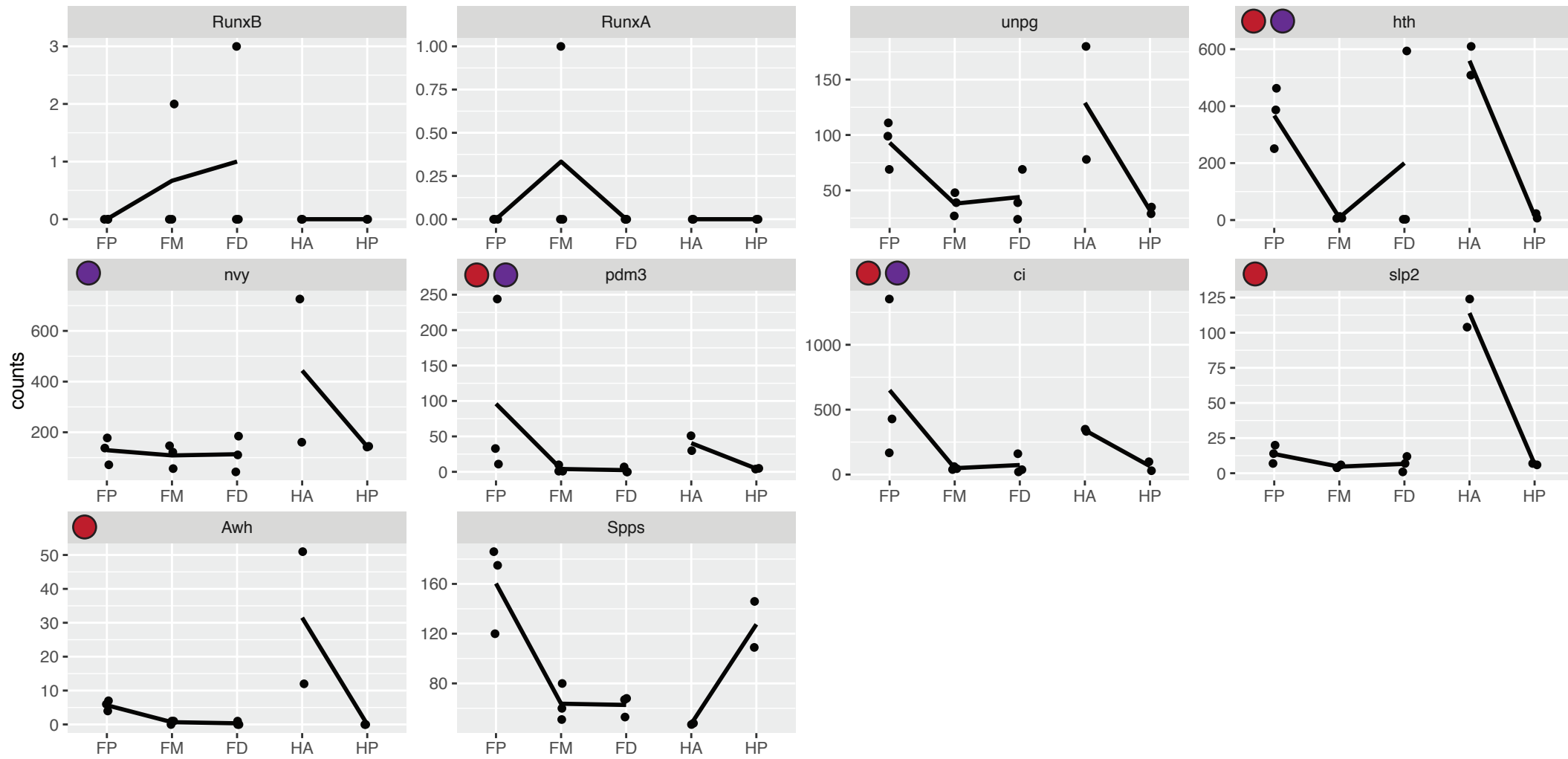


● DE in *H. mel* ● DE in *Agraulis* *H. erato* TFs Day 1 - Page 1

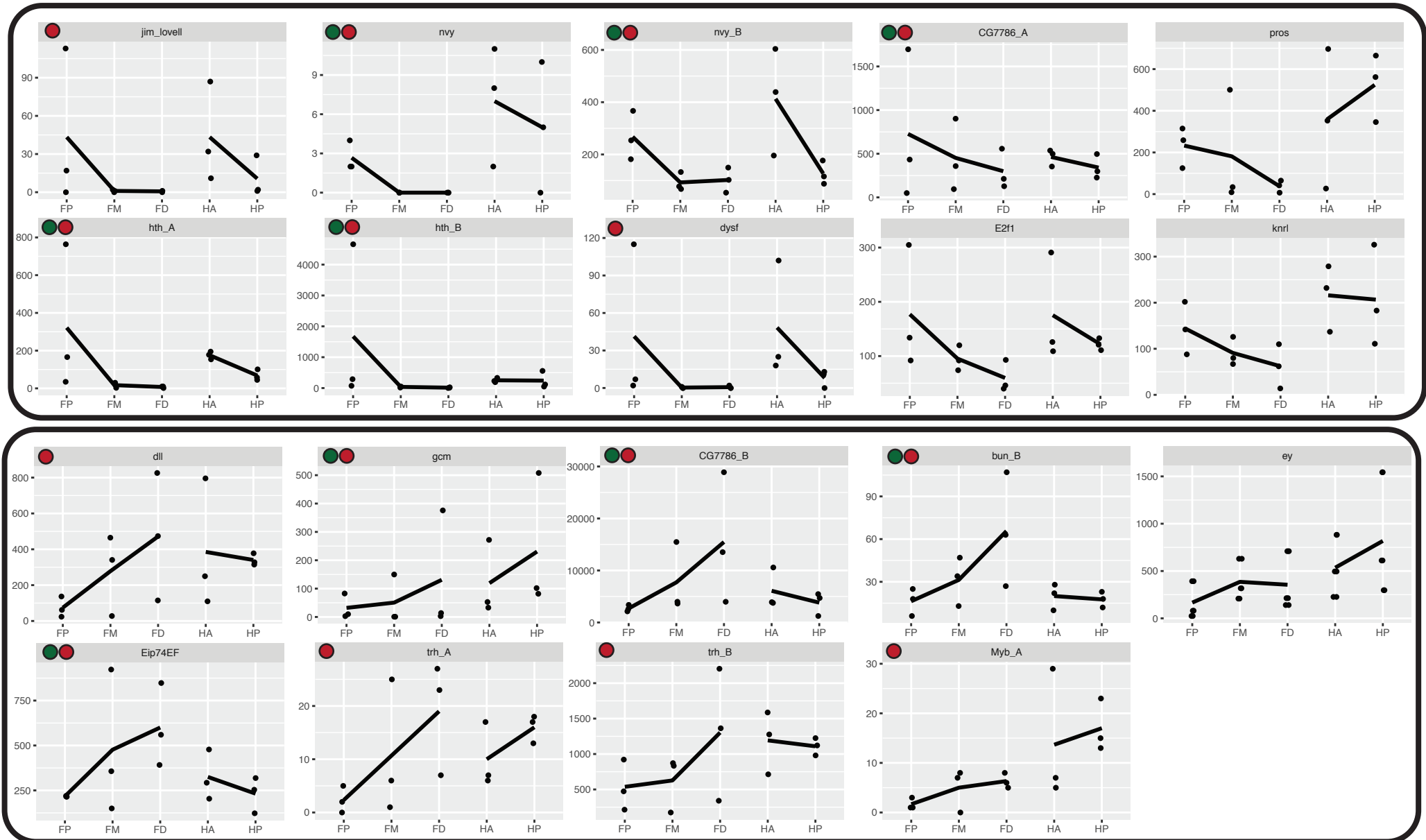




● DE in *H. mel* ● DE in *Agraulis* *H. erato* TFs, Day 2



● DE in *H. erato* ● DE in *H. mel*



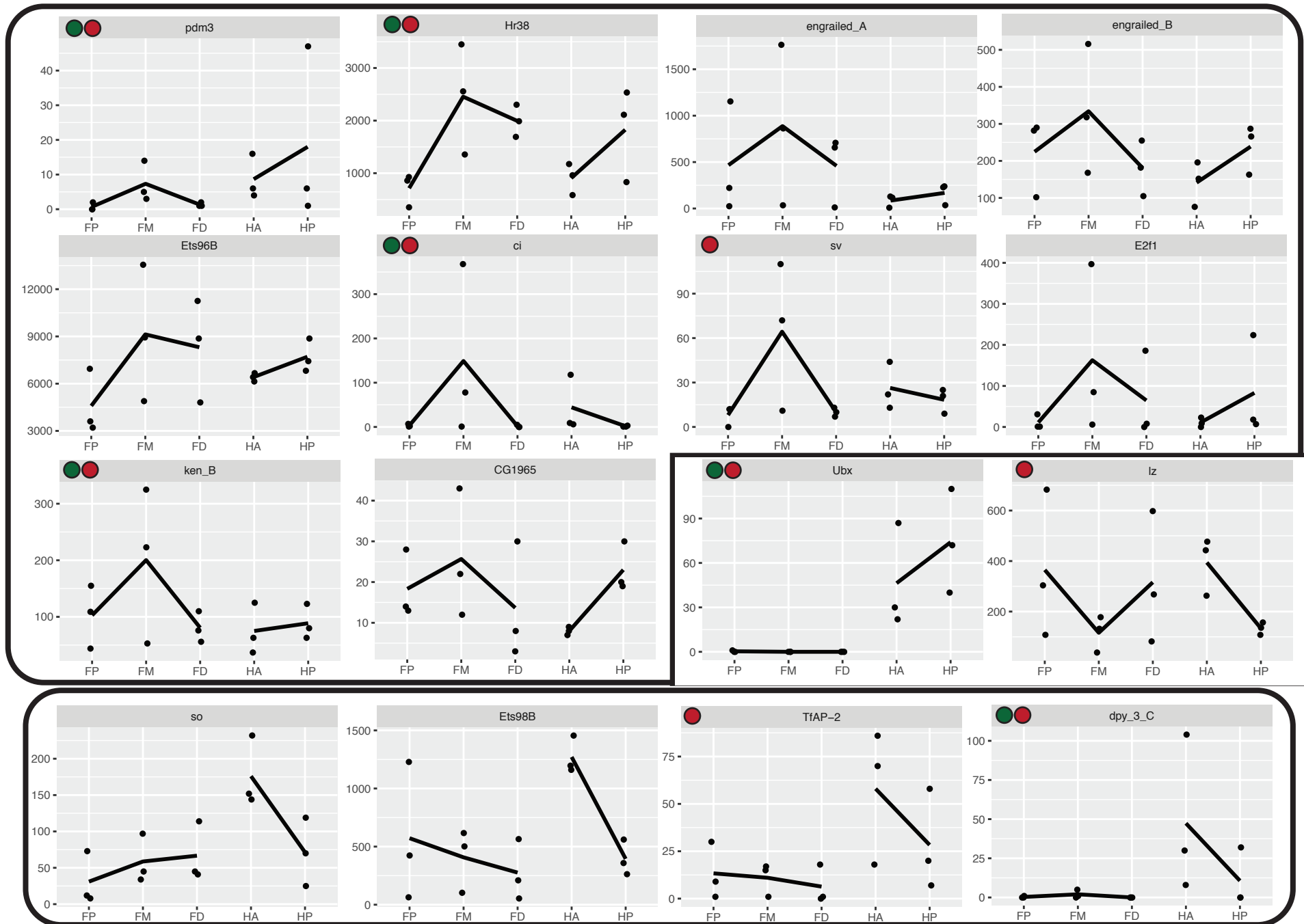
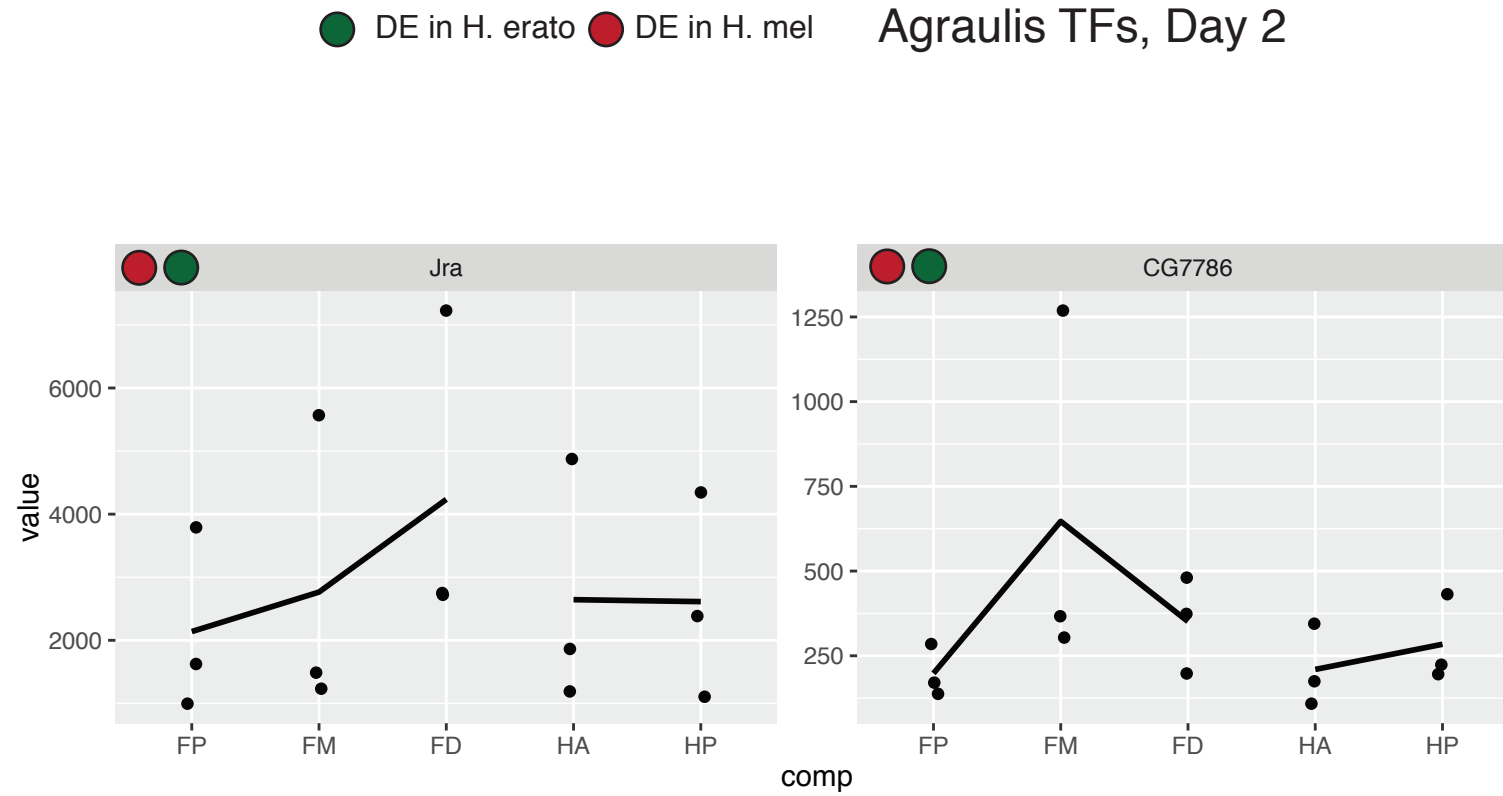


Figure 4.19



Co-differentially expressed genes

Table 4.9 lists all genes which are differentially expressed in all three species at day 1. The entry for each gene includes the expression group that the gene belongs to (Figure 4.4). At day 1, several genes involved in cuticle biosynthesis are differentially expressed in all species, as well as genes involved in wing development, neurogenesis, cytoskeletal components and signaling pathway components. Including the gene *neuralized*, a Notch-pathway E3 Ubiquitin ligase which has been identified as having a high rate of adaptive evolution between *H. erato* and *H. melpomene* (Pinharanda, unpublished).

No genes are differentially expressed in all three species at day 2, but Table 2.10 lists all genes which are differentially expressed in at least two species. In particular, a number of melanin pathway synthesis genes including *Tan* and *Ddc* are differentially expressed in both *H. melpomene* and *H. erato*.

At day 2, a higher proportion of the genes are expressed in either pattern F (low in medial forewings and higher in proximal and distal forewings) or pattern D (the opposite – high in medial forewing and lower in proximal and distal forewings). This includes the melanin pathway genes *pale* and *tan*.

Table 4.9: Genes differentially expressed in all three species – Day 1

Gene	full name	function	<i>A.v</i>	<i>H.m</i>
nov_gene015670			C	C
Hr38	Hormone receptor-like in 38	required for adult cuticle development	D	D
gcm	glial cells missing	(discussed in text)	B	B
Sse	Separase	chromatid separation in meiosis	D	D
sand	sandman	dopamine responsive potassium channel	A	A
CG2663		intracellular transport	D	D
gd	gastrulation-defective	secreted serine protease, activates Toll ligand spatzle, dorsoventral axis patterning	C	X
nov_gene011683			C	A
CG42339		immune response	C	A
GstS1	Glutathione S transferase S1		A	A
CGstD1	Glutathione S transferase D1		C	C
CG4914		serine endopeptidase activity	D	A
CG7896	no annotated func		A	D
Scp2	Sarcoplasmic calcium-binding protein		A	A
disco-r	disco-related	Epithelial-mesenchymal transition, wing expansion	E	E
sob	sister of odd and bowl	leg disc – joint morphogenesis	E	E
Reck	Reversion-inducing-cysteine-rich protein with kazal motifs	no annotated func	A	A
nov_gene021747			A	A
Faa	Fumarylacetoace tase	aromatic amino acid metabolism	C	C
Thor	Thor	translation initiation factor 4E binding protein	A	C
CG14257	no annotated func		A	A
spo	spook	CNS development, cuticle development, head involuion	A	A
CG42346	no annotated func		E	E

nov_gene007135			C	C
nov_gene007136			XX	D
nov_gene046315			B	C
Cyp6a2	Cytochrome P450-6a2	response to DDT, caffeine.	A	D
zfh2	Zn finger homeodomain 2	putative TF, proximal-distal patterning during wing and leg imaginal disc development	A	A
CG4945		mesoderm development	E	E
knrl	knirps-like	orphan nuclear hormone receptor, target of Hh, Wnt and Notch pathways	A	A
hth	homothorax		A	A
nov_gene015709			E	E
Cht2	Chitinase 2		D	D
nov_gene008541			F	C
Roc1a	Regulator of cullins 1a	E3 Ubq ligase component	C	D
CG9380			B	D
CG41378			A	E
Rap2l	Ras-associated protein 2-like		C	C
scrambl1	scramblase 1	scramble phospholipids across bilayer. (role in neurotransmission, apoptosis)	C	F
nov_gene031883			C	D
nov_gene031884			F	
pnut	peanut	component of septin GTPase complex, involved in organization of cell cortex	C	D
alpha-Catr	α -catenin related	actin binding	F	F
nov_gene			A	C
CG9701	beta-glucosidase		B	F
CG11370	no annotated func		A	A
CG15239	no annotated func		A	A
sqd	squid	RNA binding protein, localisation	E	E
Nlg3	Neurologin 3	synaptic adhesion molecule involved in synapse formation and synaptic transmission	E	E
Cyp6a18			A	E
CG15497	no annotated func		D	C
E2f1	E2F		D	B

	transcription factor 1			
Archease		axon regeneration	C	C
DCTN3-p24	Dynactin 3, p24 subunit	dynein activation	C	F
Fip1	Factor interacting with poly(A) polymerase 1		B	D
pum	pumilio	3'UTR binding protein	D	D
pum			D	D
mago	mago nashi	splicing, photoreceptor differentiation	C	D
RpS8	Ribosomal protein S8		C	D
Spn85F	Serpin 85F	chitin modifier	A	C
Mctp	Multiple C2 domain and transmembrane region protein	Ca ²⁺ ion-binding, membrane component	F	B
KFase	Kynurenine formamidase	possible pigment synthesis!	C	D
obst-A	obstructor-A	chitin maturation	E	C
obst-A			E	C
obst-B			A	A
betaTub56D	β-Tubulin at 56D	microtubule unit	E	C
CG3655	no annotated func		A	A
CG4678		metallocarboxypeptidase activity	Z	F
CG34461		chitin structural component	A	E
nov_gene034193			A	A
Lcp65Ac	Lcp65Ac	chitin structural component	A	A
Cpr49Aa	Cuticular protein 49Aa	chitin structural component	E	A
nov_gene034334			A	E
unc-5	unc-5	axon guidance	D	A
Nc73EF	Neural conserved at 73EF			D
nov_gene036907				D
Spn42Da	Serpin 42Da		C	B
nov_gene046015			C	E
beat-IV	beat-IV	heterophilic cell-cell adhesion	C	A
nGstE6			C	A
CG42259		haemolymph coagulation	F	C
CG42259			A	A

nov_gene007986			A	C
CG1402	no annotated func		C	A
Treh	Trehalase	trehalose metabooism, neural stem cell maintenance	E	B
Chd64		juvenile hormone- responsive, actin binding	D	A
CG4213	no annotated func		D	A
H2.0	Homeodomain protein 2.0		B	E
CG16885	no annotated func			E
IP3K2	IP3-kinase 2	regulates calcium levels by influencing IP3 signaling	A	B
nov_gene011713				F
CG16786	no annotated func		B	C
Bx	Beadex	dimerizes with LIM transcription factors including Apterous in the wing disc.		B
Ets98B		oocyte-expressed	C	D
neur	neuralized	endocytosis-dependent activation of Notch	XX	D
CG42390	no annotated func		A	E
Nrg	Neuroglian	axon growth, imaginal disc morphogenesis	A	A
CG40160		serine-type endopeptidase	A	C
CG8369		wing disc dorsal-ventral patterning (via interaction with Ap-Bx)	A	C
CG10407	no annotated func		A	D
Toll-7		axon guidance, viral immunity	C	A
mspo	M-spondin	regulation of myoblast fusion		E
nov_gene039499				F
ds	dachsous	cadherin binding, Planar Cell Polarity	F	C
beat-IIIb		heterophilic cell-cell adhesion		E
Nos	Nitric oxide synthase	nervous system development	E	C
TwIE	TweedleE	chitin strucutural component	E	A
CG14964		actin binding protein		E
Prm	Paramyosin	structural component of muscle	D	E

Table 4.10: Genes which are differentially expressed in at least two species at day 2.

gene	full name	function	<i>A.v</i>	<i>H.e</i>	<i>H.m</i>
DE in <i>Agraulis</i> & <i>H. erato</i>					
CG13367			E	E	
nov_gene007134			E	E	
Dnai2	dynein, axonemal, intermediate chain 2		F	E	
nov_gene00837			D	D	
nov_gene008541			F	F	
ple	pale	tyrosine hydroxylase (rate limiting step in melanin synthesis)	B	C	
Cpr65Ec	Cuticular protein 65Ec	chitin structural component	E	E	
CG10555		transcription coactivator	A	F	
Ddc	Dopa decarboxylase	melanin synthesis	B	C	
CG33290	no annotated func		B	B	
Prm	Paramyosin	structural component of muscle	E	E	
Mlc1	Myosin alkali light chain 1	myosin component	E	E	
CG9297	no annotated func		C	C	
CG11825	no annotated func		F	F	
DE in <i>H. melpomene</i> & <i>H. erato</i>					
nov_gene011805				F	F
mab-21	malformed abdomen 21	no annotated func		F	B
Sulf1	Sulfated	heparan sulfate modifier, regulates Wnt ligand diffusion, als hh signalling		A	F
gd	gastrulation- defective	secreted serine protease, activates Toll ligand spatzle, dorsoventral axis patterning		F	B
t	tan	converts NBAD to dopamine – melanin synthesis pathway		F	F
CG4757		carboxylic ester		F	B

		hydroxylase		
CG5973		transporter activity	C	C
AP-2alpha	Adaptor Protein complex 2, α subunit	cell transport	C	F
DE in <i>H. melpomene</i> & <i>Agraulis</i>				
CG33290	no annotated func		B	D
CG31954		serine endopeptidase	D	F
nov_gene011684			D	D
CG5112	no annotated func		F	F
Mctp	Multiple C2 domain and transmembrane region protein		B	C
nov_gene037141			B	E
CG13868	no annotated func		B	F
nov_gene046277			B	B
nov_gene007986			B	B

Genes with no homologs

Several genes which were identified as differentially expressed in multiple species in this experiment are labeled “nov_gene####”, indicating that they did not score any hits when compared by BLAST to the FlyBase database, were not automatically annotated with information about homology during the generation of *H. melpomene* or *H. erato* genomes, and have not been manually annotated. Some attempts can be made to identify homologs to these proteins, or at least functional domains in the coding sequence, including checking the protein against LepBase homologs, InterProtScan, and by searching with the NCBI protein domain search tool. Some of the novel genes were found to have non-arthropod homologs, others have known domains, but several continue to have no annotated functions and no known domains.

Gene ID	Function
Nov_gene008541	PAT1 super family topoisomerase
Nov_gene007135	Pupal cuticle protein PCP52
Nov_gene007136	Cuticle protein 1
Nov_gene007986	Pupal cuticle protein PCP52
Nov_gene011683	No_data
Nov_gene015670	No_data
Nov_gene015709	No_data
Nov_gene021747	contains DUF2369 super family domain, fibronectin type II domain
Nov_gene031883	homology to vertebrate mitochondrial GTPase Era, 50s ribosome binding GTPase, bacterial Ferrous iron transport protein B (FeoB) domain.
Nov_gene031884	No_data
Nov_gene034193	No_data
Nov_gene034334	No_data
Nov_gene036907	No_data
Nov_gene039499	No_data
Nov_gene046015	No_data
Nov_gene046315	No_data

IV: DISCUSSION

Despite decades of studies of the developmental diversity of butterfly wing patterns, we remain surprisingly ignorant of the full spectrum of gene networks that underlie the spatial arrangement of patterns on a butterfly wing. Here I have described these networks in a phylogenetic context, providing important background for future experimental studies. The results corroborate earlier work on differential expression of a limited set of genes and proteins in wings of other species of *Lepidoptera* (Ferguson and Jiggins, 2009, Reed and Gilbert, 2004, Weatherbee et al., 1999, Keys et al., 1999, Galant et al., 1998, Warren et al., 1994). Broadly, *H. melpomene*, *H. erato* and *Agraulis* share a common spatiotemporal transcriptomic landscape in the developing wing with other *Lepidoptera* and with *Drosophila*, implying the existence of a shared insect wing gene regulatory network. However, the results also highlight the fact that expression patterns can also evolve rapidly and in some cases are differentially expressed between the wings of these three species. In particular, the Wnt signalling pathway constituents vary widely in their expression, in ways that may be correlated with wing pattern. This data set provides a list of candidate genes for interacting with the development of wing pattern in heliconiine butterflies, and will contribute to the picture of gene expression in the wings of butterflies generally.

Construction of bioinformatic resources for Agraulis

Of the two different assemblies constructed for *Agraulis*, the Ragout/RATT annotation has generated a high-quality, stringent annotation of loci and homologous genes at the expense of genes that could not be transferred, or for which there was missing sequence in the DISCOVAR assembly, whereas the Trinity/Transdecoder transcriptome includes a much larger number of transcripts, but these transcripts are more likely to be poorly annotated and perhaps misassembled. The Trinity/TransDecoder transcriptome is likely to mis-annotate non-coding portions of transcripts such as 5' or 3' UTRs, as indicated by the discrepancy between the annotated lengths of the non-coding portion of the annotation for the gene *Ubx*, which is reflected in the difference in mapped counts. Ideally, these two approaches could be combined by using the RNAseq data to generate a *de novo* annotation of the *Agraulis* genome, but this was beyond the scope of this study and I therefore decided to use the

more complete Trinity/TransDecoder transcriptome. Additionally, I was able to identify many differentially expressed orthologues with this annotated transcriptome despite the lack of positional homology.

Weaker differential expression at day 2

Clustering by PCA shows that at day 1, samples cluster broadly by both wing and compartment, however, at day 2 the samples cluster more by individual. There are multiple explanations for this; first, some form of technical sampling error specific to day 2 could have been replicated in all species. This seems unlikely as samples were processed in controlled temperature conditions and dissected at the same time of day. Alternatively, it could mean that the amount of variability across the forewing proximodistal axis and the hindwing anteroposterior axis is less prevalent at day 2 in all three species. As most developmental processes in the wing have ceased by this point, and the majority of the remaining developmental time is given over to growth, chitin deposition and scale cell development, it is possible that fewer spatially distinct factors are now expressed, leading to a reduction in compartment specific clustering and allowing inter-individual variation to swamp the clustering analysis. “Individual” was included as a factor in the GLM, so inter-individual variation is accounted for in the statistical analysis – and is reflected in the low number of genes scored in the analysis at day 2 in all samples.

Wnt pathway components vary between species

Variance in *WntA* expression in correlation with wing pattern has previously been shown in many butterfly clades, including between races and species of *Heliconius*, and in *Agraulis* (Martin and Reed, 2014, Martin et al., 2012). In this study, I found that other Wnt pathway constituents also vary in their expression domains between species.

In particular, the Wnt pathway constituents in *H. erato* were mainly expressed in a correlated pattern – high in the proximal forewing, and low everywhere else. This pattern closely mirrors the expression profile of *WntA* in larval wing discs of *H. erato demophoon* from Panama (Martin et al., 2012). The *WntA* expression profile for Panamanian *H. melpomene rosina* is notably different from that of its co-mimic;

expression is present in the distal forewing as well as the proximal forewing, and the boundary of proximal *WntA* expression does not correlate well with the proximal boundary of the red pattern element in the adult wing. The pathway constituents are much less unified in their expression profile than those in *H. erato*, and include a number of other Wnt ligands, including *Wnt6*, *Wnt7*, and *Wg*, all of which appear to be expressed in patterns which are different from both *WntA* and each other.

Recent implementation of the CRISPR-Cas9 system in multiple *Heliconius* species has allowed functional testing of *WntA*. In *H. erato* *WntA*^{-/-} G₀ knockouts, the red band expands proximally down to the base of the forewing while the distal boundary remains unchanged (Martin et al, in submission). Additionally, heparin-injected *H. erato* lose their forewing band pattern elements. In contrast, expression and functional evidence from *H. melpomene* suggests that *WntA* plays a different role. This difference is reflected in *H. melpomene* *WntA*^{-/-} G₀ CRISPR knockouts, in which the proximal boundary is only slightly perturbed, including the appearance of a yellow spot. This provides an indication that while these factors are not the targets of selection at the within-species level of divergence, regulatory diversification of the Wnt pathway has occurred between *Heliconius* species, implying that this aspect of the wing gene regulatory network has been dynamic in recent evolutionary history (Figure 4.20). This implies that developmental morphospace has been fundamentally altered in each lineage. Alternatively, it is possible that the coordinated expression of Wnt-pathway constituents in pupal wings is shaped or reinforced by earlier expression of *WntA* in larval wing discs, and that there has been no direct change to the regulation of other factors. This could be tested by expression analysis of these Wnt pathway constituents in the *WntA*^{-/-} G₀ CRISPR knockouts.

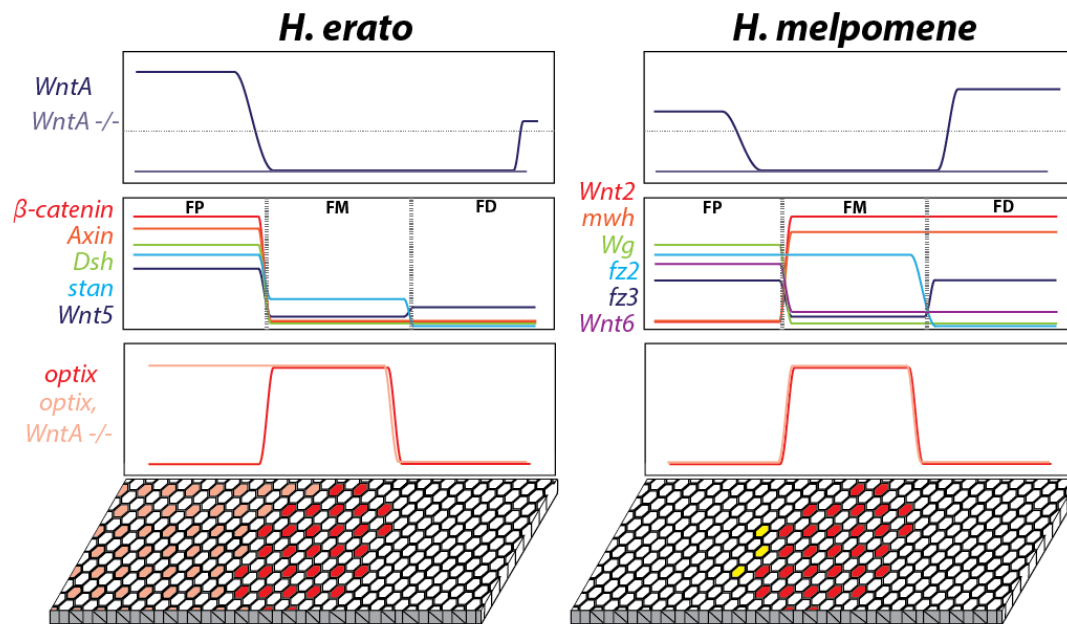


Figure 4.20: A recapitulation of chapter 1 figure 6. The top panel shows the wild-type expression profile for *WntA* in *H. erato* and *H. melpomene* as inferred from Martin et al (2012). The bottom panel indicates the resultant expression profiles of *optix* in these mutants as determined by Wallbank (unpublished) and the resultant aberrant wing patterns. We can now add additional Wnt-pathway factors into this view of the development of wing pattern. Some of these factors may be capable of influencing the signal cascade that regulates *optix* in *H. erato*, and it is possible that the boundaries of the red band are demarcated by a different Wnt pathway component in *H. melpomene* – though it is also possible that this is achieved by a different mechanism entirely.

Observation of the role of Wnt-pathway constituents in *Agraulis*, a related species with a wing pattern that more closely resembles the “classical” Nymphalid patterns, can inform our understanding of which scenario is likely. Fewer Wnt-pathway elements were detected as differentially expressed in *Agraulis* – many of the factors that were differentially expressed were co-differentially expressed in *H. erato*, but generally the expression profiles of these factors was not restricted to the strong proximal expression profile seen in *H. erato*. Additionally, the *Agraulis WntA* expression domain does not correlate with expression of Wnt factors observed here. Generally, there is poor correlation between expression of Wnt pathway constituents in the three species. This seems to support regulatory divergence between these lineages.

In *Agraulis*, the consequences of *WntA* manipulation are dependent on wing compartment. Heparin-injected butterflies show an expansion of silver elements in the posterior hindwing and loss of these elements in the forewing, whereas *WntA*^{-/-} G₀ CRISPR knockouts show the opposite effect, with expansion of silver elements in the anterior hindwing and reduction in the posterior hindwing. A number of elements which are differentially expressed between the anterior and posterior hindwing could participate in this, including the gene *Wntless*, which is upregulated in anterior hindwings and is required for Wnt ligand secretion, and could therefore significantly alter the dynamics of Wnt signalling in the two compartments.

Transcription factors

Transcription factors provide the physical interactions that lead to differential regulation in gene regulatory networks. Their expression domains are therefore critical to understanding how the wing develops, and to how this process generates diverse wing patterns.

Several transcription factors that are known to be involved in development of wings in *Drosophila* and *Junonia* were identified in this experiment in their expected expression profiles, including *Ubx*, *Ci* and *hth*. *Ubx* specifies appendage identity on the T3 segment, and here was found to be expressed in hindwings. *Ci* is expressed in the anterior compartment, which requires *smo* and *ci* to be competent to respond to Hedgehog signalling, and here, *ci* was expressed in the anterior compartment of the

hindwing. Homothorax is a homeodomain transcription factor, and in *Drosophila* patterns the haltere and wing hinge. Expression of homothorax protein has been assayed by immunohistochemistry in *H. melpomene*, and is shown to be expressed in the proximal third of the wing. The recapitulation of these expression profiles in all three species serves as validation for the experiment.

Several other factors were consistently differentially expressed in all three species. These factors could constitute the developmental morphospace along the forewing proximodistal axis and hindwing anteroposterior axis in pupal wings. They include *jra*, the transcription factor downstream of the JNK-signalling pathway – particularly notable as the JNK pathway receptor *domeless* has been identified as a candidate for involvement in yellow pattern elements.

Optix was not identified as differentially expressed in this data set, in any of the three species at any stage. However, in *H. melpomene*, a non-coding transcript near *optix*, which I manually annotated after the recognition of non-coding transcripts at the *cortex/domeless/washout* locus, is upregulated in the medial forewing – i.e. in the pattern that *optix* would be expected to be expressed in, based on *in situ* hybridisation and immunohistochemistry. Immunohistochemical and *in situ* hybridisation against *optix* may be better able to detect low levels of expression than RNAseq. Also the strongest period of expression of *optix* is later than the time points in this experiment. It is possible that the differential expression of this non-coding transcript precedes differential expression of the *optix* transcript itself, which could be further investigated.

Several transcription factors were found to be differentially expressed in two or three species in this study. Of these, many, but not all, were expressed in conserved expression profiles. Some, for instance *bunched*, had a conserved expression pattern between *H. melpomene* and *H. erato*, but a different expression pattern in *H. erato*. In addition to this, a number of transcription factors were identified in each species that were private to that species. So while I have been able to identify additional elements of the conserved gene regulatory network that are expressed in the wing of butterflies, there are also elements that vary between species, and that are absent from others. This assists in building a picture of how wing patterns might develop in *Heliconius*,

but also how the transcriptomic landscape in the wing might evolve in butterflies more generally (Figure 4.21).

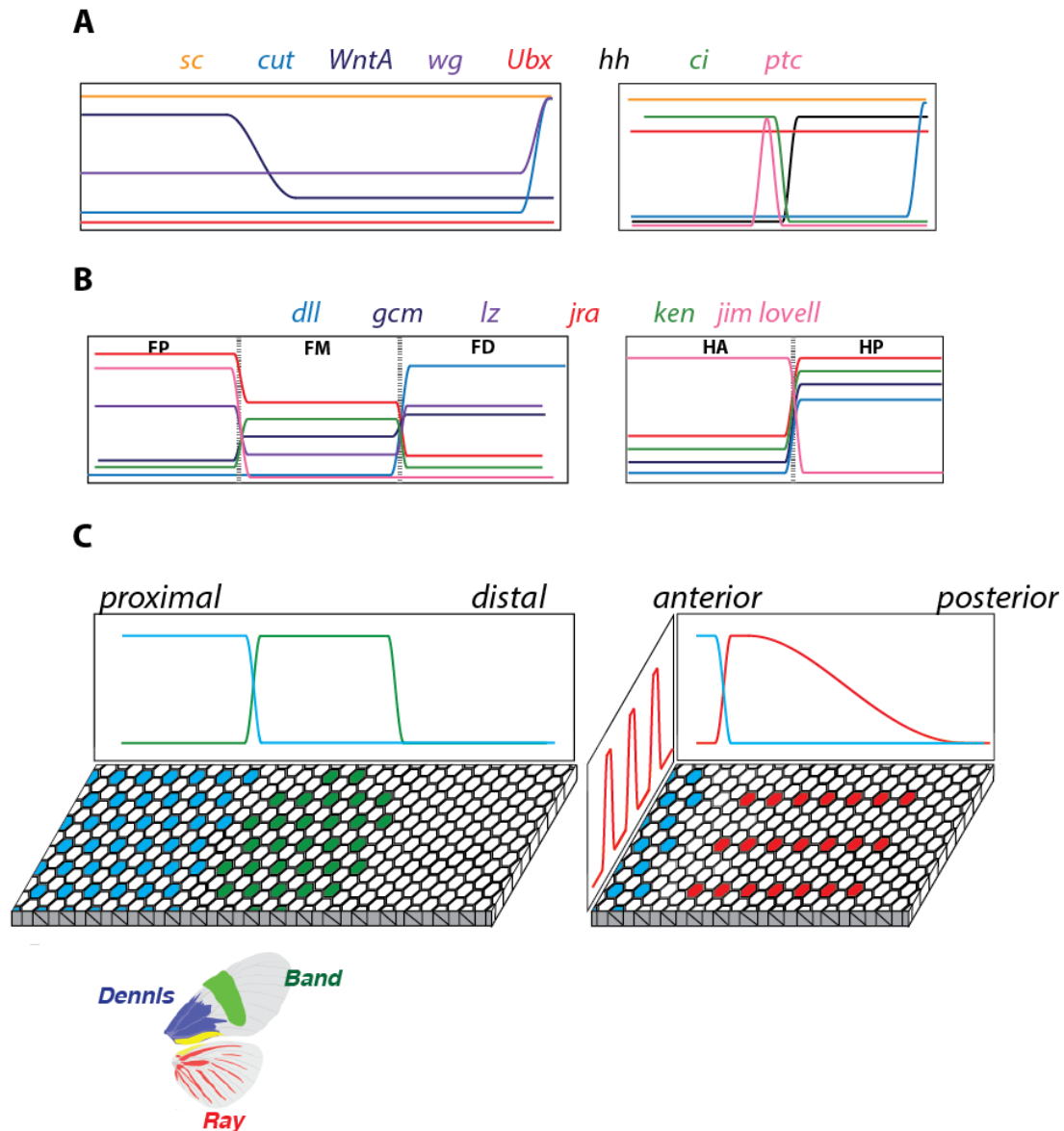


Figure 4.21: Cartoon summary of transcription factor expression domains across the proximodistal axis of the forewing and the anteroposterior axis of the hindwing. Panel A illustrates the known expression domains of factors in wings of *Junonia* and other butterflies listed in chapter 1 and depicted in chapter 1 figure 5. Panel B shows a selection of transcription factor expression patterns described in this experiment, including both factors known to pattern wings in other insects (eg *Dll*), as well as factors that have not previously been described as expressed in wings (e.g. *jim lovell*, a transcription factor associated with gravity sensation in *Drosophila*). Panel C shows the red band pattern elements.

Other co-differentially expressed genes

Other than Wnt pathway constituents and transcription factors, a number of other genes were found to be differentially expressed which could affect the development of wing pattern. Notably, many chitin deposition-related factors were found to be differentially expressed in all three species, as well as cytoskeleton-interacting factors and β -tubulin. The expression of such genes in developing wings might be expected to occur evenly across the wing, particularly during the cellular morphogenesis of scale cells, which require large cytoskeletal movements for laminar extension, and chitin deposition factors for cuticle synthesis. It is known that scale cells of different types show developmental heterochrony; type I (yellow and white scales) begin to develop earlier than other scales, as illustrated by Aymone with electron microscopy (Aymone et al., 2014). It is possible that this process has begun at the level of gene expression before laminar extension begins (when the differentiation becomes visible with microscopy). If type I scales were already developing at this time, it would imply that the window of competency to affect the differentiation of type I scales may have ended before day 1 of pupal development. Alternatively there might just be some timing differences across the wing – for example maybe chitin deposition begins proximally, or just unevenly across the wing, or it could even be related to vein density.

A large number of genes relating to neural development are differentially expressed across wings. This is perhaps related to the fact that lepidopteran wing scales are homologous to bristle cells in other insects. Bristle cells are a sensory structure; the initial precursor cell of the bristle undergoes two rounds of division, the first creating a neural precursor and the second the bristle precursor, and the second division splitting the neural precursor into a neuron and a glial cell, and the bristle precursor into a socket cell and the bristle-proper. In lepidopteran scales, the same cell fate lineage is followed, but the neuron and glial precursor undergoes apoptosis.

Finally, several melanin-pathway genes are observed to be differentially expressed across wings, including *Ddc*, *pale* and *tan*. Previous work has focused on the differential expression of pigment synthesis pathway genes in *Heliconius*, in particular looking for evidence of convergent evolution expression domains of genes

in the melanin and ommochrome synthesis pathway in correlation with mimetic pattern elements in *H. erato* and *H. melpomene* (Ferguson and Jiggins, 2009, Ferguson et al., 2011). The result here, showing differential expression of genes involved in the melanin synthesis pathway but not genes from the ommochrome pathway, matches the findings by Ferguson et al that some melanin pathway genes have an early phase of expression in early post-pupation stage wings (excluding the gene *yellow*, also not detected here), becoming differentially expressed again in the final stages of pigment deposition, whereas ommochrome pathway genes do not have an early phase of expression, only being expressed at later stages of wing development.

Building a model for wing pattern development and evolution

In summary, there is a set of genes that are differentially expressed in the wings of these heliconiines, as well as in *Junonia* and *Drosophila*. I propose that these constitute a conserved wing gene regulatory network, and that these genes are candidates for affecting the development of wing pattern. This has been achieved without a candidate gene-driven approach, which has allowed the identification of factors that may participate in wing patterning that are lineage specific, or which may vary significantly in their expression profile between lineages.

Under Nijhout's interpretation of the Nymphalid ground plan, we might have predicted very high conservation of gene expression with only subtle changes of vein-related factors. Under Gilbert's windows and shutters model, one would have expected to see *Heliconius*-specific modifications that would explain the propensity for diversity in their wing patterns. The reality is a combination of both of these. There are clearly a number of deeply-conserved expression domains along the proximo-distal axis of the forewing and the anteroposterior axis of the hindwing. At the same time, the Wnt pathway has clearly diversified within each lineage. Given the pre-existing evidence of conserved gene expression in *Junonia*, and of different effects of Wnt perturbation in different species, perhaps this result is not surprising.

Conclusion

I have described the expression of genes in the developing wings of three species of butterfly. I detected differential expression of key developmental factors involved in determining the identity, axes and morphology of insect wings generally. Many of these factors, along with others not previously described as being expressed in developing wings, were expressed in conserved domains of expression in all three species. Other factors, most notably the constituents of the Wnt signalling pathway, varied in different lineages with wing pattern. Together these genes constitute the gene regulatory network which is present in the wing during the specification of wing pattern. A deeper understanding of factors that are expressed in the wing in correlation with pattern elements will assist in decoding the regulatory linkages that lead to the differential expression of the switch genes *optix*, *WntA* and *cortex*, and will hopefully be of use to the general understanding of butterfly wing pattern evolution.

A key issue for evolutionary developmental biology is to resolve how natural diversity can form in the face of developmental robustness. By understanding how pattern and form have evolved in the context of an adaptive radiation, we can make inferences about how the processes of evolution may sculpt developmental processes more generally. In this thesis, I have investigated how modification and conservation of aspects of a gene regulatory network have influenced evolution of form in different lineages, first at the level of *cis*-regulatory modification, then at the level of patterning and switch genes, and finally at the level of conserved developmental networks. Each of these studies contributes to our understanding of how gene regulatory networks can be modified to create diversity.

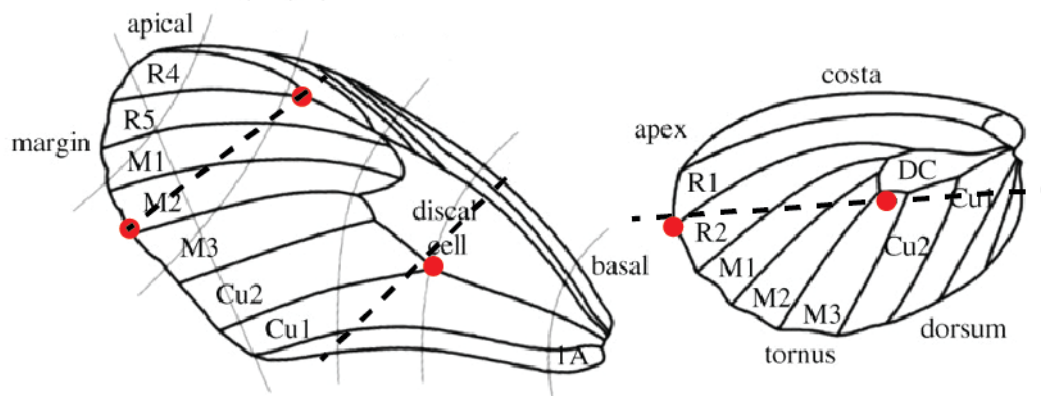
Supplementary table 4.1: number of reads per sample, % of reads which failed to map in each sample.

<i>H. melpomene</i>		
Sample	Read count	% reads that do not map
13F_FP1	10342965	10.90
13G_FM1	14953963	9.56
13H_FD1	10205562	9.99
13I_HA1	11397631	13.31
13J_HPo1	12267336	11.93
14A_FP1	14892190	14.16
14B_FM1	12295189	12.44
14C_FD1	11528254	14.92
14D_HA1	10717086	15.46
14E_HPo1	15162459	12.37
14F_FP1	12040114	13.71
14G_FM1	12411856	12.78
14H_FD1	12303691	11.12
24D_HA1	13131538	32.69
14J_HPo1	13945427	14.40
15A_FP2.	11437808	13.87
15B_FM2.	10993529	12.81
15C_FD2.	11856974	13.56
15D_HA2.	10581808	13.95
15E_HPo2	13681765	11.02
16A_FP2.	14736312	8.84
16B_FM2.	13626298	9.40
16D_HA2.	12686225	4.96
16E_HPo2	12823781	9.95
16H_FD2.	15598458	10.46
17F_FP2.	11015917	7.64
17G_FM2.	12136669	7.46
17H_FD2.	14589976	7.52
17I_HA2.	14663720	7.54
17J_HPo2	14600239	7.81

<i>H. erato</i>		
Sample	Read count	% reads that do not map
B8_FP1	13286804	7.56
C3_FD1	15104587	9.84
C6_HA1	12841432	6.31
D10_FP1	15919485	8.05
F21_HPo1	14646293	8.35
F8_FM1	14965698	10.61
G21_FM1	13046272	8.03
G5_FM1	12543875	7.66
H4_HA1	13419952	7.91
I8_FD1	14022680	10.02
S3_HPo1	24574848	6.98
A4_HPo2	12912895	18.97
B2_FD2	13435820	19.43
B4_HA2	12182847	18.86
C4_FM2	14127156	20.68
C9_FP2	15100384	19.49
D6_HA2	14027838	17.72
D9_HPo2	12354324	17.87
E3_FM2	13926825	23.33
E6_FP2	13622452	15.73
G6_FD2	13281517	19.52
J4_FD1	14498546	16.59
J8_FP2	16882308	19.12
N3_FM2	13139640	18.78

<i>Agraulis vanillae</i>			
Sample	Read count	% reads that do not map, Trinity	% reads that do not map, Ragout
11F	21601472	10.78	17.85
11G	19505732	10.39	16.90
11H	9472553	11.65	19.17
11I	20695225	16.69	20.40
11J	15779259	12.80	24.99
12A	17909757	18.40	19.70
12B	21362299	9.44	21.08
12C	19020085	10.37	21.63
12D	21698478	11.32	19.69
12E	21574847	11.73	20.14
12F	19511174	10.86	20.93
12G	23853328	11.20	21.92
12H	23137900	10.96	20.01
12I	18912480	12.82	21.79
12J	18012267	13.16	20.32
21A	25105152	11.28	17.85
21D	21967288	11.08	16.90
21E	22224755	12.23	19.17
22G	18790997	7.53	20.40
22H	19159884	7.12	24.99
23A	16849671	8.19	19.70
23B	22852458	7.84	21.08
23C	20814817	7.90	21.63
23D	21198748	7.27	19.69
23E	21347562	7.28	20.14
23F	22977885	8.09	20.93
23G	19254249	7.79	21.92
23H	19779195	7.88	20.01
23I	20971110	8.80	21.79
23J	23057419	7.51	20.32

Figure S4.1 Wing dissection diagram. Pupal forewings were cut into proximal, medial and distal sections; the cuts were chosen to surround the red forewing band. The key landmarks are highlighted in red. These are the junction between the discal cell and the Cu1 vein, for the proximal section, the junction between the R4 and R5 veins, and the point at which vein M2 meets the wing margin for the medial/distal sections. Pupal hindwings were cut into approximate Anterior and Posterior pieces; the cut was chosen to keep the region of the wing that forms the yellow band entirely on one side of the cut. While no pattern is visible yet in pupal wings, the venation pattern can clearly be seen. The key landmarks used are highlighted with red circles; these were the posterior base of the discal cell (DC), and the point where the R1 vein contacts the wing edge.



References

- ALTSCHUL, S. F., MADDEN, T. L., SCHAFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-402.
- AYMONE, A. C., LOTHHAMMER, N., VALENTE, V. L. & DE ARAUJO, A. M. 2014. Embryogenesis of *Heliconius erato* (Lepidoptera, Nymphalidae): a contribution to the anatomical development of an evo-devo model organism. *Dev Growth Differ*, 56, 448-59.
- BENSON, W. W. 1971. Evidence for the evolution of unpalatability through kin selection in the Heliconinae (Lepidoptera). *The American Naturalist*, 105, 213-226.
- BRAUS, H. 1900. Die Muskeln der Nerven Ceratodusflosse Ein Beitrag zur vergleichenden Morphologie der freien Gliedmaße bei niederen Fischen und zur Archipterygiumtheorie. *Denkschr. d. med.-nat. Ges. Jena* 4, 137-300.
- BRISCOE, A. D., MACIAS-MUNOZ, A., KOZAK, K. M., WALTERS, J. R., YUAN, F., JAMIE, G. A., MARTIN, S. H., DASMAHAPATRA, K. K., FERGUSON, L. C., MALLET, J., JACQUIN-JOLY, E. & JIGGINS, C. D. 2013. Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genet*, 9, e1003620.
- BROWN, K. S. & MIELKE, O. H. H. 1972. Heliconians of Brazil (Lepidoptera - Nymphalidae) .2. Introduction and General Comments, with a Supplementary Revision of Tribe. *Zoologica*, 57, 1-&.
- CARROLL, S. B. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell*, 101, 577-80.
- DAVEY, J. W., CHOUTEAU, M., BARKER, S. L., MAROJA, L., BAXTER, S. W., SIMPSON, F., JORON, M., MALLET, J., DASMAHAPATRA, K. K. & JIGGINS, C. D. 2016. Major Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3 (Bethesda)*, 6, 695-708.
- FERGUSON, L. C. & JIGGINS, C. D. 2009. Shared and divergent expression domains on mimetic *Heliconius* wings. *Evol Dev*, 11, 498-512.
- FERGUSON, L. C., MAROJA, L. & JIGGINS, C. D. 2011. Convergent, modular expression of ebony and tan in the mimetic wing patterns of *Heliconius* butterflies. *Dev Genes Evol*, 221, 297-308.

- GALANT, R., SKEATH, J. B., PADDOCK, S., LEWIS, D. L. & CARROLL, S. B. 1998. Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. *Curr Biol*, 8, 807-13.
- GILBERT, L. 2003. Adaptive novelty through introgression in *Heliconius* wing patterns: evidence for shared genetic "tool box" from synthetic hybrid zones and a theory of diversification. *Ecology and evolution taking flight: butterflies as model systems*, 281-318.
- GOMPEL, N. & PRUD'HOMME, B. 2009. The causes of repeated genetic evolution. *Dev Biol*, 332, 36-47.
- GRAMATES, L. S., MARYGOLD, S. J., SANTOS, G. D., URBANO, J. M., ANTONAZZO, G., MATTHEWS, B. B., REY, A. J., TABONE, C. J., CROSBY, M. A., EMMERT, D. B., FALLS, K., GOODMAN, J. L., HU, Y., PONTING, L., SCHROEDER, A. J., STRELETS, V. B., THURMOND, J., ZHOU, P. & THE FLYBASE, C. 2017. FlyBase at 25: looking to the future. *Nucleic Acids Res*, 45, D663-D671.
- HAAS, B. J., PAPANICOLAOU, A., YASSOUR, M., GRABHERR, M., BLOOD, P. D., BOWDEN, J., COUGER, M. B., ECCLES, D., LI, B., LIEBER, M., MACMANES, M. D., OTT, M., ORVIS, J., POCHET, N., STROZZI, F., WEEKS, N., WESTERMAN, R., WILLIAM, T., DEWEY, C. N., HENSCHER, R., LEDUC, R. D., FRIEDMAN, N. & REGEV, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, 8, 1494-512.
- HOFF, K. J., LANGE, S., LOMSADZE, A., BORODOVSKY, M. & STANKE, M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32, 767-9.
- JIGGINS, C. D., WALLBANK, R. W. & HANLY, J. J. 2017. Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? *Philos Trans R Soc Lond B Biol Sci*, 372.
- JONES, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S. Y., LOPEZ, R. & HUNTER, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236-40.
- KEYS, D. N., LEWIS, D. L., SELEGUE, J. E., PEARSON, B. J., GOODRICH, L. V., JOHNSON, R. L., GATES, J., SCOTT, M. P. & CARROLL, S. B. 1999. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science*, 283, 532-4.

- KOCH, P. B. & NIJHOUT, H. F. 2002. The role of wing veins in colour pattern development in the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae). *European Journal of Entomology*, 99, 67-72.
- KOLMOGOROV, M., RANEY, B., PATEN, B. & PHAM, S. 2014. Ragout-a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*, 30, i302-9.
- KOZAK, K. M., WAHLBERG, N., NEILD, A. F., DASMAHAPATRA, K. K., MALLET, J. & JIGGINS, C. D. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic heliconius butterflies. *Syst Biol*, 64, 505-24.
- LOMSADZE, A., BURNS, P. D. & BORODOVSKY, M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res*, 42, e119.
- LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15, 550.
- MARTIN, A., PAPA, R., NADEAU, N. J., HILL, R. I., COUNTERMAN, B. A., HALDER, G., JIGGINS, C. D., KRONFORST, M. R., LONG, A. D., MCMILLAN, W. O. & REED, R. D. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proc Natl Acad Sci U S A*, 109, 12632-7.
- MARTIN, A. & REED, R. D. 2014. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Dev Biol*, 395, 367-78.
- NIJHOUT, H. F. 1991. The development and evolution of butterfly wing patterns. *Smithsonian series in comparative evolutionary biology (USA)*.
- NIJHOUT, H. F. & WRAY, G. A. 1988. Homologies in the colour patterns of the genus *Heliconius* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 33, 345-365.
- NIJHOUT, H. F., WRAY, G. A. & GILBERT, L. E. 1990. An analysis of the phenotypic effects of certain colour pattern genes in *Heliconius* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 40, 357-372.
- PATEL, N. H., MARTIN-BLANCO, E., COLEMAN, K. G., POOLE, S. J., ELLIS, M. C., KORNBERG, T. B. & GOODMAN, C. S. 1989. Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell*, 58, 955-68.
- PIERCE, S. E., HUTCHINSON, J. R. & CLACK, J. A. 2013. Historical perspectives on the evolution of tetrapodomorph movement. *Integr Comp Biol*, 53, 209-23.

- REED, R. D. & GILBERT, L. E. 2004. Wing venation and Distal-less expression in *Heliconius* butterfly wing pattern development. *Dev Genes Evol*, 214, 628-34.
- RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16, 276-7.
- SALAZAR, C., BAXTER, S. W., PARDO-DIAZ, C., WU, G., SURRIDGE, A., LINARES, M., BERMINGHAM, E. & JIGGINS, C. D. 2010. Genetic evidence for hybrid trait speciation in *heliconius* butterflies. *PLoS Genet*, 6, e1000930.
- SCHWANWITSCH, B. 1924. 21. On the Ground-plan of Wing-pattern in Nymphalids and certain other Families of the Rhopaloeerous Lepidoptera. *Journal of Zoology*, 94, 509-528.
- SHEPPARD, P., TURNER, J., BROWN, K., BENSON, W. & SINGER, M. 1985. Genetics and the evolution of Muellierian mimicry in *Heliconius* butterflies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 433-610.
- STANKE, M. & MORGENSTERN, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*, 33, W465-7.
- STERN, D. L. & ORGOGOZO, V. 2009. Is genetic evolution predictable? *Science*, 323, 746-51.
- SUFFERT, F. 1927. Zur Vergleichende Analyse der Schmetterlingszeichnung. *Biologisches Zentralblatt*, 47, 385-413.
- VAN BELLEGHEM, S. M., RASTAS, P., PAPANICOLAOU, A., MARTIN, S. H., ARIAS, C. F., SUPPLE, M. A., HANLY, J. J., MALLET, J., LEWIS, J. J., HINES, H. M., RUIZ, M., SALAZAR, C., LINARES, M., MOREIRA, G. R. P., JIGGINS, C. D., COUNTERMAN, B. A., MCMILLAN, W. O. & PAPA, R. 2017. Complex modular architecture around a simple toolkit of wing pattern genes. *Nat Ecol Evol*, 1.
- VAN SCHOOTEN, B., JIGGINS, C. D., BRISCOE, A. D. & PAPA, R. 2016. Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. *BMC Genomics*, 17, 254.
- WALTERS, J. R., HARDCASTLE, T. J. & JIGGINS, C. D. 2015. Sex Chromosome Dosage Compensation in *Heliconius* Butterflies: Global yet Still Incomplete? *Genome Biol Evol*, 7, 2545-59.

- WARREN, R. W., NAGY, L., SELEGUE, J., GATES, J. & CARROLL, S. 1994. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature*, 372, 458-61.
- WEATHERBEE, S. D., NIJHOUT, H. F., GRUNERT, L. W., HALDER, G., GALANT, R., SELEGUE, J. & CARROLL, S. 1999. Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr Biol*, 9, 109-15.
- WERNER, T., KOSHIKAWA, S., WILLIAMS, T. M. & CARROLL, S. B. 2010. Generation of a novel wing colour pattern by the Wingless morphogen. *Nature*, 464, 1143-8.
- YU, Q. Y., FANG, S. M., ZHANG, Z. & JIGGINS, C. D. 2016. The transcriptome response of *Heliconius melpomene* larvae to a novel host plant. *Mol Ecol*, 25, 4850-65.